# The Explanatory Structure of Moral Worth

Harjit Bhogal

**Abstract**

This paper is about the nature of *accidentality* and what that tells us about which actions are *morally worthy*.

The central dispute about moral worth is between those who think that morally worthy actions are motivated by the *fact that the action is right* and those who think they are motivated by *right-making features* of an action. A variety of authors, from Kant onwards, have argued for the former view – since we could be motivated by the features that make an action right but still do the right thing merely accidentally.

This paper investigates the broader concept of accidentality as it applies in various domain across philosophy. Particularly relevant are considerations originating in the philosophy of science about the nature of *coincidence*. As a result of this investigation I formulate, and defend, a novel account of moral worth that is based on the idea that there should be a *unified explanation* of why the agent did the right thing. Such a view is a powerful argument against views which say worthy actions must be motivated by the rightness of the action itself.

Some actions are morally right. That is, sometimes there is a matching or correspondence between the action performed and the facts about rightness – $\phi$ was performed and $\phi$ is right.

Of these right actions, only some have *moral worth*. (I'm assuming, as is common in the literature, that morally worthy actions are objectively morally right.)[1] The politician who does charity work solely for publicity [Sliwa, 2016, p. 393] and the person who acts randomly but luckily don't act worthily. They don't deserve credit for their right actions.

These examples, along with many others in the literature, suggest that morally worthy actions must be done for the right reasons. There are two main approaches to which reasons are right:

**Rightness Itself (RI)**: Morally worthy actions are motivated by the fact that the action is right

**Rightness Making Features (RMF)**: Morally worthy actions are motivated by the features of the action that make it right[2]

These are slogans, not precise formulations, but they point to the central division in the literature. Arpaly [2002] and Markovits [2010], among others, hold RMF views – a worthy action is motivated by the fact it would reduce suffering, or would keep a promise to your friend, or so on. Others, in the spirit of Kant, [e.g. Herman, 1993, Sliwa, 2016, Johnson King, 2020] claim that a worthy action is motivated by the fact that it is right.

[1] Markovits [2010] and Howard [2021] are, in somewhat different ways, exceptions.

[2] These approaches go by various names. This terminology is from Singh [2020].

However, RMF views face a major problem. It seems that an agent could be motivated by right-making features, yet only *accidentally* do the right thing. But an accidentally right action isn't worthy.

Consider, for example, Singh's [2020] **Venom** case (which is representative of a class of cases):

> **Venom** Jack, a surgeon, is hiking when he sees a stranger get bitten by a venomous snake and faint. He immediately makes an incision near the bite so that the venom will drain out. Making the incision is the right thing to do, and Jack's reason for doing it (that it will allow the venom to drain out) is part of what makes it right. But Jack doesn't have any particular concern for doing the right thing in this case, nor does he conceive of his reason as one that makes his action right. He is simply intrinsically interested in draining venom out of wounds. (p. 162)

Jack is motivated by a right-making feature. Nevertheless, his making the incision isn't morally worthy. The problem, it seems, is that the RMF view allows us to be insufficiently attentive to rightness so it can *just so happen* that our motivation lines up with what makes the action right.

RI views, on the other hand, seem to avoid accidentality worries – if an action is motivated by rightness then it's not accidentally right. But we will see that things are more complicated than that.

In this paper I will investigate the nature of accidentality as it's relevant to whether an action is worthy. (In fact, my focus initially will be on *coincidentality*, I'll discuss that distinction very soon.) Recent developments in the moral worth literature introduce a variety of concepts that bear upon accidentality in different ways.[3] But the core of accidentality is rather unclear.

A guiding thought is that the concepts of flukes, accidents and coincidences that are central to the moral worth literature aren't unique to that domain. So considering how these concepts work in other domains will shed light on moral worth. Particularly relevant will be considerations originating in the philosophy of science about the nature of coincidence.

I'll argue for two things. First, that accidentality and coincidentality should be understood in explanatory terms, as opposed to, for example, modal terms. This affects how we should understand both RI and RMF views.

Second, understanding the nature of coincidence lets us formulate a very attractive version of the RMF view. It does better than existing RMF accounts at avoiding accidentality while retaining the core intuitions that motivate RMF approaches. In fact, it avoids accidentality just as well as RI views.

This is a powerful argument against RI views, since their main motivation is ruling out accidentally right actions. Markovits [2010, p. 206], for example, notes that (Kant's version of) the RI view 'gained what attraction it held from the plausibility of the thought that morally worthy actions don't just happen to conform to the moral law – as a matter of mere accident.' Influential arguments for RI views, from Kant's shopkeeper to the present day, are driven by an effort to rule out accidentally right actions. If my RMF view avoids accidentality as successfully as RI views then it becomes hard to see why we would favor an RI view.

[3] For example, Isserow [2019, section 4] on *competent causation*, Cunningham [2021], Lord [2017] on *manifesting know-how*, Way [2017] on *moral principles*, Singh [2020] on *the guise of the good*, Johnson King [2020] on *deliberateness*, and so on.

In section 1 I'll discuss the relation between accident and coincidence. In section 2 I'll consider various approaches to coincidence – arguing that explanatory approaches are preferable. I section 3 I'll discuss how to formulate the RI view in light of this. In sections 4-6 I'll formulate the RMF view in light of this. In sections 7 and 8 I'll discuss how this RMF view deals with commonly discussed cases and how it compares to previous accounts in the literature. In section 9 I'll discuss, again, the relation between accident and coincidence, focusing on a concern about deviant explanatory chains. Section 10 concludes.

## 1 Accidents and Coincidences

The concept *accident* is part of a cluster of concepts that aren't typically distinguished in the literature on moral worth. It will be helpful to focus on one of these concepts: *coincidence*.

Many kinds of things are called accidents. The concept of coincidence is more tightly circumscribed. The literature on coincidence suggests that there are two parts to a coincidence. Firstly, there are two or more component events that match in a striking way.[4] Secondly, those matching component events are, in some sense, not properly related or connected (e.g. Hart and Honoré [1985, p. 74], Lando [2017], Bhogal [2020], Baras [2020], Berry [2020]).

Some examples: I toss a fair coin twenty times and it lands heads every time. There is a striking match between the component coin tosses but the tosses seem unrelated. So, it's a coincidence that they all landed heads. You end up on a cruise with your old enemy [Owens, 1992], so there is a striking match between component events – your location and your old enemy's location. If those events are not properly connected then it's a coincidence. There are coincidence problems in cosmology – for example there is a striking match between the amount of energy in the universe coming from dark matter and the amount that is dark energy, but these quantities don't seem connected [Bhogal, 2020, p.677-8]. And debunking arguments against moral realism often claim that the striking match between our moral beliefs and the moral truth – that we have true moral beliefs – would be a coincidence if (certain forms of) moral realism were true since because our beliefs and the truth are would not be properly related [Field, 1996, Street, 2006, among many others].

Questions about moral worth fit this structure. Jack, in **Venom**, makes the incision. Making the incision was right. There is a striking match between the action performed and the facts about rightness. But are those facts connected in the appropriate way? If they are not, and Jack only coincidentally did the right thing, then his action is not worthy. For much of the paper, then, I'll focus on coincidence and how to rule out coincidentally right actions.

However, the exact connection between *coincidence* and *accident* is complicated. Later, I'll discuss cases where it's not a coincidence that the agent does the right thing but perhaps it's still an accident.

## 2 Approaches to Coincidence

So, let's start with coincidence. Again, a coincidence is a striking match between component events that are not properly related. But what is it to be 'properly' related? What relation dispels coincidence? Let's consider two

[4] See Baras for a comprehensive recent discussion of strikingness. Luckily, giving an account of strikingness won't be necessary here.

prima facie plausible approaches.

## 2.1 The Modal Approach

The first approach is modal – a striking match between events A and B is non-coincidental if certain modal conditions hold. Take the twenty coin tosses – it could *very* easily have been the case that not all landed heads. If I had spun the coin a fraction more on the fourteenth toss, for example, it would have landed tails.

So, perhaps, a matching between events is non-coincidental if it could not easily have failed to hold. This condition on non-coincidence is closely related to *safety* conditions designed to rule out luckily true beliefs. Roughly, my belief in a proposition is safe if I could not easily have falsely believed that proposition.

However, this view faces the problem of *modally robust coincidences*. Consider mathematical coincidences. It's a coincidence that the numbers 31, 331, 3331, 33331, 333331, 3333331 and 33333331 are each prime – 333333331 is not prime Lange [2010]. But this coincidence could not easily have failed to hold – it is necessary.

Similarly, we can imagine coincidences that hold with the necessity of the laws of nature. Consider this case from Bhogal (fort.):

> **Protons and Electrons** Protons are positively charged. Electrons are negatively charged. However, the absolute value of their charge is the same. Specifically, protons have a charge of $1.602176634 \times 10^{-19}$ coulombs, while electrons have a charge of $-1.602176634 \times 10^{-19}$ coulombs.

If it were a basic law of nature that protons have that charge, and separately, a basic law of nature that electrons have that charge then the matching of charge would be nomically necessary. But, still, if that was all that could be said then it would be a coincidence that the charges matched.

So, modal robustness isn't, on it's own, enough to dispel coincidence. Maybe the solution is to add another modal condition. We have been considering a condition closely related to *safety* conditions on knowledge or justification. This suggests adding a condition in the spirit of *sensitivity*. Roughly, my belief in a proposition is sensitive if had the proposition been false then I would not have believed it [Nozick, 1981].

But this condition doesn't help with the cases just considered. In fact, it's somewhat hard to even make sense of. If 31 hadn't been prime would 331, 3331 and 33331 have not been prime? There seems to be nothing substantive to say here. Similarly, if it wasn't a law that electrons have a charge of $1.602176634 \times 10^{-19}$ coulombs would it still be a law that protons have a charge of $1.602176634 \times 10^{-19}$ coulombs? It doesn't seem reasonable to rest the distinction between coincidence and non-coincidence on such questions. (And, as we will see in section 3.1, sensitivity-style conditions have additional problems in distinguishing worthy from non-worthy actions.)

There is a lot more to say about the complicated literature on safety and sensitivity-like conditions in necessary domains. But this would get us too far off track. It is more useful to, now we have noted that robust coincidences make the modal approach to coincidence look uncompelling, stress that modal criteria seem particularly unsuitable for developing an account of moral worth.

### 2.1.1 The Pertinence Constraint

Some views of moral worth are modal, in the sense that the moral worth of an action depends upon what the agent would have done in other circumstances. But, as is often noted, such views seem to inappropriately conflate the moral worth of an action with a broader judgment of the agent's character. The action of a fanatical dog-lover who risked her life to save strangers is still worthy even if, were her dog in danger, she would have saved the dog instead [Markovits, 2010, p. 210]. That the dog-lover would have saved the dog is bears upon her character but does not warrant withholding credit for this action. Similarly, Sliwa [2016, p. 399-400] claims that 'when deciding whether to give an agent credit for an action…we are interested in the motivations that in fact led the agent to act.' Isserow [2019] calls this idea the *pertinence constraint* – only the motives that actually led to action are relevant for moral worth, not counterfactual ones.

Importantly, the pertinence constraint allows that counterfactuals can be *evidentially* relevant – how an agent would act in other cases can tell us about their actual motives. But it tells us that moral worth isn't directly *determined* by such counterfactuals. Consequently, it rules out using modal approaches to coincidence as part of an account of moral worth.

The pertinence constraint arises from sharply separating the moral worth of an action from a broader evaluation of the agent's character. But there is a more far-reaching and ambitious way to view it. We can see it as part of a broader 'postmodal' approach to philosophy – an approach that has been influential in recent metaphysics. One core thread of the postmodal approach is that the key metaphysical questions are not about modal facts, rather, they are questions about the structure and nature of the actual world. Modal facts are 'are often epiphenomenal, a mere reflection of deeper postmodal structure' [Sider, 2020, p. 3]. So physicalism, for example, shouldn't be understood as a claim about supervenience – this modal fact is a mere symptom of actual world facts, perhaps about what grounds what or about which metaphysical laws hold (see, for example, Kim [1993, p. 167], Schaffer [2009, p. 364], Wilsch [2017, section 3] expressing such ideas).

It's natural to extend this approach to ethics and epistemology. What matters for what you should do; what you should believe; what you know; and so on, is the structure of the actual world. Facts about other possible situations are not directly relevant. It's the actual situation that matters. Determining whether this postmodal thought can be developed in a thoroughgoing way is a vast project, far beyond anything we can discuss here. But the pertinence constraint can be seen as part of this postmodal approach – as a commitment to grounding moral evaluation in the actual.

* * *

Cases of robust coincidences give us reason to reject the modal approach to coincidence and the pertinence constraint gives us reason to think that it shouldn't, in any case, be applied in accounts of moral worth.

## 2.2 The Explanatory Approach

A more attractive approach – in line with the literature on coincidence – is to take coincidence to be an explanatory notion. The rough idea is that in a coincidence the component events are explanatorily 'disconnected' [Owens, 1992, Lando, 2017, Bhogal, 2020]. The twenty coin tosses that all landed heads seem explanatorily independent – that is why the sequence was a coincidence.

The explanatory approach has no problem with modally robust coincidences – there can be a modally robust matching between events that are nevertheless explanatorily disconnected. Consider **Protons and Electrons**. The theory that posits separate basic laws governing the charge of the proton and the electron implies both that the charges are explanatorily disconnected and that the matching between them is nomically necessary.[5]

Further, an explanatory approach to moral worth doesn't violate the pertinence constraint. The actual explanatory connection between action and rightness matters – other possible situations aren't directly relevant.[6]

Giving an account of 'explanatory disconnection' is a difficult task – one that we will come back to in section 4. But a broadly explanatory approach to coincidence is plausible.

* * *

One, perhaps obvious, point before we move on. The coincidence-dispelling relation can't be merely correlational.That is, we can't show some matching between events to be non-coincidental by just pointing to other matchings or correlations. We can't show a matching between fact A and B to be non-coincidental by pointing to a matching between A-type facts and B-type facts more generally, or by pointing to a matching between A and C. The problem, of course, is that these latter correlations could themselves be coincidental. Correlations or matchings cannot, on their own, dispel coincidence.[7]

## 3 Formulating the RI account

This discussion of coincidence has implications for moral worth. Both RI and RMF accounts need to rule out coincidentally right actions. Let's focus, in this section, on how RI accounts can do that.

The RI account's slogan is that morally worthy action is motivated by the action's rightness. However, the phrase 'motivated by the action's rightness' is ambiguous. On one natural reading for an agent to be motivated by $\phi$'s rightness is for there to be some connection between the agent's action and the actual fact of $\phi$ being right. But this is not the typical reading in the moral worth literature. The alternate, more common, reading is that an agent is motivated by $\phi$'s rightness when it's part of the content of the agent's motivational state to $\phi$ that $\phi$ is right. (One piece of evidence for this reading of the literature comes from commonly discussed cases, one of which I'll describe

[5] Applying this account to mathematical coincidences requires a story about mathematical explanation. See Mancosu [2018, sections 4-7] for a survey.

[6] The issue is made complicated by accounts of explanation where explanations themselves are determined (in part) by modal facts. The spirit of the pertinence constraint is that the actual situation is what's relevant for moral worth. But does that rule out modal facts playing *any* role in grounding the relevant features of the actual situation? This is difficult question, one which I'm not able to to take on here. But, the pertinence constraint suggests that we should reject accounts of moral worth which appeal *directly* to modal considerations and not via those modal considerations determining explanatory facts.

[7] An issue analogous to that of footnote 6 arises here. It is, perhaps, possible to have a *Humean* account of modal or explanatory facts where such facts are ultimately reduced to patterns of actual events. But we should reject accounts of coincidence that appeal directly to correlations and not via them determining facts about modal or explanatory relations.

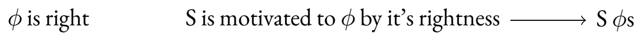

Figure 1: The correlational RI view

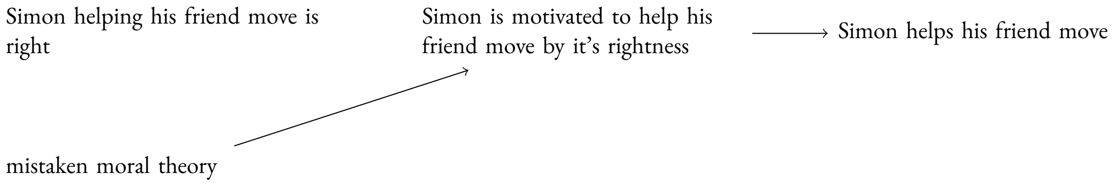
Simon helping his friend move is right
Simon is motivated to help his friend move by it's rightness
Simon helps his friend move
mistaken moral theory

Figure 2: Singh's **Moving** case

in more detail imminently, where an agent is trying to do the right thing but makes some moral mistake. However, they luckily do the right thing, so their action is not explained by the actual rightness of the action. Such cases are often thought of as problems for the slogan the an action has moral worth if it's motivated by the actions rightness – but this only makes sense on the latter, weaker, reading of what motivation consists in.[8])

I'm not meaning to commit to any view in the debate about what motivation consists in. In the bulk of the paper, though, I'll understand motivation in this latter, weaker way, since it fits with better with the moral worth literature. But this issue will come up again when we are discussing authors who may have a stronger conception of motivation.

Understood in this weaker way, the slogan that morally worthy action is motivated by the action's rightness doesn't postulate any modal or explanatory connection between an agent performing an action and the actions rightness. Rather, it merely postulates a matching between the agent's motivational state and the moral facts. As such, it's the wrong type of approach for ruling out agents coincidentally doing the right thing. Figure 1 illustrates the structure.

In the figure the arrows represent explanatory connections. There is an explanatory connection between S's motivation and their action. But there is no connection postulated between the actual fact of $\phi$'s rightness and the motivation. So, there is a matching between action and rightness, but no connection is postulated between them.

If the RI view is understood in this merely correlational way then there are counterexamples, like this one, from Singh [2020]:

**Moving** Simon's friend is in a tough spot and needs last minute help moving. Simon helps him, which is the right thing to do. And he is motivated to do so because he believes that it is the right thing to do. According to Simon's conception of morality, what is right is what benefits one's friends and harms one's enemies. But the content of his motivation is simply that helping his friend is the right thing to do.

Figure 2 illustrates this structure.

Simon holds an incorrect moral theory that, in this case, just so happens to yield the same result as the correct moral view. His action is so disconnected from the actual rightness of helping his friend that it's a coincidence that he did the right thing. (In fact, let us stipulate that this is how the case is to be understood – Simon's motivation is not explanatorily connected to the actual moral facts. It's not a case of him 'dimly seeing' the moral facts, rather it's him not seeing the moral facts at all.)

[8] See, for example, Johnson King's [2020] 'Promise Keeping' example; Singh's [2020] 'Moving' example; the 'Eichmann' case discussed by Arpaly [2002] and Sliwa [2016].

$$\begin{tikzcd}\phi \text{ is right} \arrow[r] & \text{S is motivated to } \phi \text{ by it's rightness} \arrow[r] & \text{S } \phi\text{s}\end{tikzcd}$$

Figure 3: The Explanatory RI view

A correlational version of the RI view fails, as we would expect given that mere correlations can't dispel coincidence.

### 3.1 Modal RI vs Explanatory RI

One possible fix is to postulate a modal connection between the action and it's rightness. The problem, of course, is that this builds in the modal conception of coincidence that was criticized in section 2.1.

Here's a natural way to formulate the modal RI view: S's $\phi$-ing has moral worth if and only if S is motivated by $\phi$'s rightness and S could not easily have acted wrongly with respect to $\phi$.

This view faces the problem of robust coincidences. Imagine an person who refrains from killing someone for fun because they think that refraining is right. However, they think that because they hold very strongly a deeply mistaken moral theory where they value suffering and think that people should be kept alive longer so they have more opportunities to suffer. The agent does the right thing by accident even though they could not easily have acted wrongly, since it could not easily have been the case that they killed people for fun, nor that killing people for fun was acceptable.

Adding a sensitivity-style condition doesn't help. One might say that this agent's action doesn't have moral worth because he would still refrain from killing for fun even if the moral facts were different and it was ok to kill for fun. But this doesn't distinguish this case from actions that do have moral worth.

Imagine an person who refrains from killing someone for fun because they think that refraining is right. They think that because they hold very strongly a moral theory that values human life. Plausibly, this person would not act differently even if the moral facts were different and it was ok to kill for fun. But that doesn't mean their action lacks worth.Sensitivity-style conditions don't help.

Further, modal RI approaches violate the pertinence constraint. What a person would do in other situations doesn't seem to determine the moral worth of this particular action. Recall Markovits's fanatical dog lover from section 2.1.1.

The modal RI approach is flawed because it builds in a modal account of coincidence.

An explanatory version of the RI view (figure 3) looks more promising. The slogan that morally worthy action is motivated by the action's rightness is understood as positing a explanatory connection between action and rightness. This plausibly makes it non-coincidental that the agent did the right thing.[9]

[9] Though whether a particular RI view is best understood as modal or an explanatory can be a tricky question. For example, how Sliwa's [2016] influential view should be classified plausibly depends on our background view of knowledge (see Isserow [2019, section 3.2]).

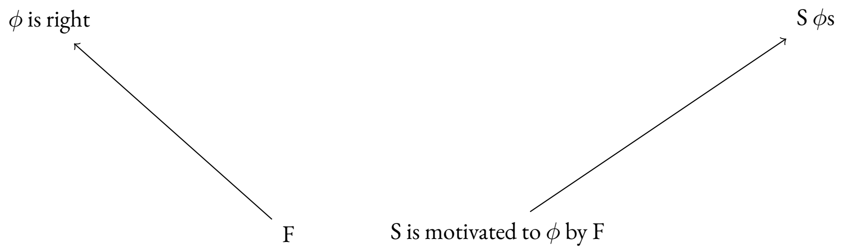

Figure 4: The Correlational RMF view

## 4 Coincidence and the RMF account

I argued that the RI account should be developed in an explanatory way since the modal approach to coincidence is flawed. The RMF view should also be developed in an explanatory way, for very similar reasons. The basic structure of an explanatory RI view is simple – the agent's action is explained, at least in part, by the rightness of the action. But it's not so easy to see what a successful explanatory RMF view – one that rules out coincidentally right actions – will look like. That's what I'll consider in this section and the next.

The RMF slogan is that morally worthy actions are motivated by the features of the action that make the action right.[10] Just as with the RI view, there is a way of understanding this which does not postulate any connection between the rightness of the action and it being performed.

Fix upon some right action $\phi$. What I'll call the *correlational RMF view* says that the agent $\phi$-ing has moral worth if (i) the action has some feature, F, that makes it right and (ii) the content of their motivation to $\phi$ is that $\phi$ has feature F. This structure is illustrated in figure 4. Again, the arrows represent explanatory connections.

Because there is no explanatory connection between F and the agent's motivation the action's rightness is explanatorily disconnected from it's being performed. This makes the correlational RMF view is subject to counterexamples. For example:

**Hiring** Stephanie is a hiring manager and gives Yi-joon a job. The content of Stephanie's motivation is that Yi-joon is the most skilled programmer. Yi-joon is the most skilled programmer and that is a good reason to give them the job. However, Stephanie only believes that Yi-joon is most skilled because of incorrect racial stereotypes.

Stephanie's action doesn't have moral worth since her motivation to give the job to Yi-joon because he is the most skilled programmer is disconnected from the actual fact of him being the most skilled programmer.

The obvious fix is to require an explanatory connection between the actual right-making feature and the agent's motivation. This structure is shown in Figure 5.

Call this the *Third-Factor RMF view*. This view posits an indirect explanatory connection between $\phi$'s rightness and the agent $\phi$-ing since there is common explainer of those facts.

[10] There is typically not just one right-making feature of your action. Often there is a vast nexus of facts relevant to making your action right [Fogal and Worsnip]. So there is a hard question about how many, and which, of the right-making features you should be motivated by, for the action to have moral worth. This question is largely orthogonal to the issues about accidents and coincidences we are considering so I leave it for other work.

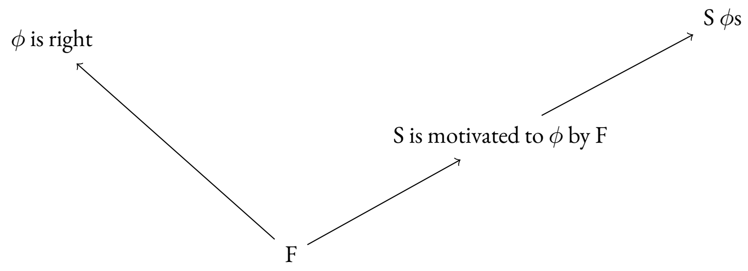

Figure 5: The Third-Factor RMF view

## 4.1 Jean

Unfortunately, such an indirect explanatory connection isn't enough – there are still coincidentality problems. Consider Sliwa's [2016] much-discussed case of Jean:

> Jean's friend missed her bus to work and frets over being late to an important meeting; coming late would be a great embarrassment to her. Wanting to spare her friend a major embarrassment, Jean gives her a ride. Let's assume that giving her friend the ride is the right thing to do in these circumstances and the fact that it spares her friend a major embarrassment makes it right. (p. 398)

Sliwa claims that 'it is a fluke that Jean did the right thing' since if Jean acts *only* out of concern for saving her friend embarrassment then what about a case where the only way to save her friend embarrassment would be to murder her friend's ex-boyfriend? Jean's doing the right thing can seem rather precarious. Consequently, Sliwa thinks, Jean's action doesn't have moral worth, even though she is motivated by a right-making feature.

From what's been said, though, it's not clear Jean's action lacks worth. There are **Bad Jean** variants of the case, for example, where Jean really would murder her friend's ex-boyfriend to save her friend embarrassment. In these cases Jean giving her friend does not seem worthy. **Bad Jean** is a counterexample to the Third-Factor RMF view.

But there are **Good Jean** variants where Jean wouldn't murder the ex-boyfriend and, we can assume, would generally act in a reasonable way in situations where her friend might face embarrassment. Here it's intuitive that Jean's act is worthy.

It might seem easy enough for the RMF theorist to deal with these cases, then. Good Jean acts with worth and Bad Jean doesn't because of the difference in how Jean would act in other situations. This is, in effect, to drop the third-factor version of the view and accept a version of the *Modal RMF* view.

But, as Sliwa [2016, p.399-400] notes these counterfactuals can't determine moral worth – that would violate the pertinence constraint and confuse the moral worth of this action with a judgment of Jean's character.

The most forceful way to understand Sliwa's argument, I think, is that because Bad Jean doesn't act with moral worth and we can't draw the distinction with Good Jean based on counterfactuals then there is no relevant distinction, and the actions of Good Jean aren't worthy either. This is a deep problem for the RMF view since Good Jean is the type of case RMF theorists identify as paradigmatically worthy. (Though this interpretation goes a little beyond Sliwa's text – whether Sliwa's intended argument is of exactly this form is not clear to me.)

Thus the RMF view faces a challenge. What's the difference between **Bad Jean** and **Good Jean**? Correlational approaches are non-starters; modal approaches fail the pertinence constraint; and the explanatory approach we have considered – the *Third-Factor RMF view* doesn't help with the Jean case. So where does the RMF theorist go from here?

## 5 'Stapling together' explanations and the anti-coincidence condition

Notice that the Correlational RMF view failed because it doesn't establish a connection between the agent's action and rightness. See figure 4 again. There is an explanation of why $\phi$ is right, and a distinct explanation of why S $\phi$-ed. If we try to explain why S did the right thing all we can do is 'staple together' these two explanations – but this isn't enough to show that the agent non-coincidentally did the right thing.

This 'anti-stapling' thought is common in debunking arguments against moral or mathematical realism: On such views we can separately explain what's morally or mathematically right, and why we have the beliefs that we do, but those are just disconnected explanations stapled together. So, it would just be a coincidence if our moral or mathematical beliefs were true. Street [2016, p. 31], for example, expresses this idea:[11]

> One may explain each side of the coincidence in as much depth as one likes – going into wonderful normative depth about why family and friendship are valuable, and wonderful scientific depth about why we were selected to think this. But all this goes nowhere toward explaining the thing that really needs to be explained, namely the coincidence itself.

Similarly, consider again the case of the 20 coin tosses that landed heads. Consider an explanation of this that separately appeals to the precise microphysical details of each coin landing heads – explaining why the first toss landed heads, then explaining why the second landed heads, and so on. This explanation isn't satisfying. It doesn't show us that the sequence was non-coincidental. All we are doing here is stapling together separate explanations.

As Field [1996] and Lange [2010], among others, note, what we want instead – what dispels coincidence – is a *unified* explanation. What we want in the coin toss case is an explanation that 'brings together' all the instances – for example, an explanation that notes that the coin was heavily weighted to heads. And what Street wants is a story about how our beliefs can be genuinely connected to the moral facts, not just a story about our beliefs conjoined with a story about the moral facts.

The moral here is that explaining why S $\phi$-ed, and why $\phi$ was right is very different from explaining *why S did the right thing*. Of course, that S $\phi$-ed and $\phi$ was right *implies* that S did the right thing. And that S did the right thing is *metaphysically explained* by S $\phi$-ing and $\phi$ being right. But still, a satisfying explanation of why S $\phi$-ed and $\phi$ was right is very different from a satisfying explanation of why S did the right thing.

A natural way to develop these thoughts (following Bhogal [2022]) is that there is a unified explanation of why the agent did the right thing when the *best explanation* of why the agent did the right thing isn't merely an explanation

[11] See also Field [1996, section 5], Linnebo [2006, section 2], Berry [2020, section 5b].

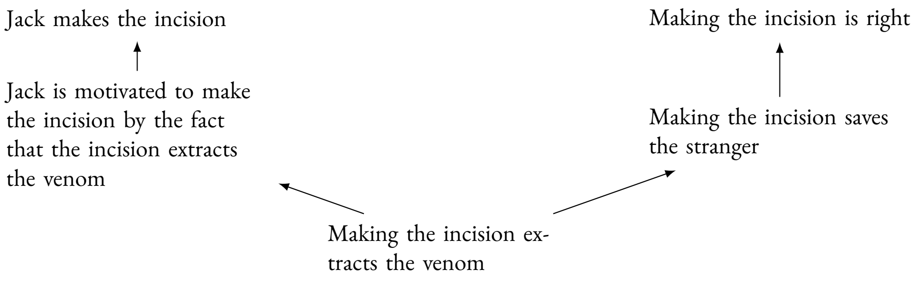
Jack makes the incision
Jack is motivated to make the incision by the fact that the incision extracts the venom
Making the incision is right
Making the incision saves the stranger
Making the incision extracts the venom

Figure 6: Venom

of the agent doing $\phi$ and $\phi$ being right. This condition is a very direct way of ruling out explanations that are merely 'stapled together'. For the agent to non-coincidentally do the right thing we need a unified explanation – an explanation of why the agent did the right thing that is not merely an explanation of why they did the thing and why that thing was right.

This is just an outline of an account of unified explanation – one that I think is fairly minimal and perhaps even uncontroversial. To fill out this account we would need to add a full theory about what explanations are good or bad in various domains. But this outline is enough for our purposes in the rest of the paper.

Notice, though, that this condition implies that there can be a common explainer of S doing $\phi$ and $\phi$ being right but it is still coincidental that S did the right thing.

For example, in **Venom** there is a common explainer of Jack making the incision and making the incision being right, but if we try to explain why Jack did the right thing all we can do is explain why Jack made the incision – because of his intrinsic interest in extracting venom – and why it was right to make the incision – because it would save the victim's life. All we can do, that is, is explain why agent did $\phi$ and why $\phi$ was right. There is no unified explanation. Consequently, it's a coincidence that Jack did the right thing. Figure 6 illustrates this. In fact, this is the source of counterexamples to third-factor RMF view. The indirect explanatory connection between rightness and action that the third-factor view postulates isn't enough – we need a unified explanation.[12].

Now we are in a position to see what an RMF account has to look like to avoid coincidentality worries.

**Unified Explanation RMF:** For an agent's doing $\phi$ to have moral worth (i) the agent must be motivated by right-making features of $\phi$ and (ii) there must be a unified explanation of why the agent did the right thing – that is, the best explanation of why the agent did the right thing cannot be just an explanation of why the agent did $\phi$ and why $\phi$ was right.

This account is very simple – condition (i) is what makes it an RMF view; condition (ii) is the anti-coincidence condition.

Here is the plan for the rest of the paper: The central motivation for this view so far involved considering the general nature of coincidence. In the next few sections I'll compare this view to others in the moral worth literature, as well as seeing how it deals with some key cases. In section 6 I'll compare it to RI views. In section 7 I'll consider what the

[12] This is in the spirit of certain arguments made in the literature on debunking – that an indirect explanatory connection between our moral beliefs and the moral truth, a third-factor, isn't enough to show that the correlation between moral belief and moral truth is unproblematic [Korman and Locke, 2020, Lutz, 2018, Faraci, 2019, Bhogal, 2022]

view says about some commonly discussed cases. In section 8 I'll compare it to some RMF views in the literature. Finally, notice that this view, as it stands, states a necessary condition for moral worth. We will discuss sufficiency in section 9.

## 6 Unified Explanations and Rightness Itself

For RMF accounts to avoid coincidentality they must take on part of the spirit of the RI approach. If an agent does right thing non-coincidentally there is a *unified* explanation of why the agent *did the right thing*, not just an explanation of why the agent did $\phi$ and why $\phi$ was right. But this is to say that rightness itself, and not just the details of the right action, have an explanatory role to play.

But Unified Explanation RMF doesn't collapse into an RI view. For an agent's action to be non-coincidentally right there has to be an explanatory connection between rightness itself and the action. But this doesn't imply that the agent must be motivated by rightness itself.

How might rightness itself play an explanatory role, and an agent act non-coincidentally rightly, without the agent being motivated by rightness itself? I'll just mention a few possibilities:

(1) Perhaps the agent is motivated directly by right-making features but has a kind of background attentiveness to rightness itself which acts as a 'filter' 'plac[ing] limits upon an agent's capacity…to act on other motives' [Isserow, 2021, p. 281] so that motives that would lead to immoral actions are filtered out.[13] This would, for example, prevent Jean from acting out of the motive of saving her friend embarrassment when that motive would lead to murdering the ex-boyfriend. This filter would generate a relevant explanatory connection between rightness and the agent's action.

(2) Perhaps the agent is motivated directly by the right-making features of an action, but the reason that the agent is motivated by those features is that they have some direct insight into the moral facts. Such an agent might just form a desire to save their friend from embarrassment, for example. But, ultimately, that desire is formed because of some insight they have into rightness, even if they don't conceptualize it as such, even if they don't think of their action as morally right, and even if they don't think of their action in moral terms at all. This insight – the connection the agent has to the actual facts about rightness – would provide a unified explanation for why the agent did the right thing.

(3) Similarly to the last option, perhaps the agent is motivated directly by the right-making features of an action, but the reason that the agent is motivated by those features is because they have gone through some reliable process of moral education. The agent may have been brought up to be be kind and generous and save their friends from embarrassment where possible. Because their moral education is reliable the actions of this agent will be connected to facts about rightness. Presumably, the reason that the agent had a reliable moral education involves, at some point in the causal chain, a connection to rightness – perhaps on the part of their teachers, or their teachers' teachers. So, there is a relevant connection between rightness and the agent's action.

[13] Stratton-Lake [2000] is plausibly understood as requiring that such a filter is present for an act to have worth.

In fact, this is a very common case. An agent acts generously because of their moral character, and this moral character is due to their moral upbringing. But an agent need not be motivated by rightness itself, or even think of their action in moral terms, when they are acting.

(1), (2) and (3) illustrate ways in which there could be a unified explanation of the agent doing the right thing without the agent being motivated by rightness itself. In particular, notice that (2) and (3) involve a connection between the agent's action and rightness but they do not require that the agent know, or even believe, that what they are doing is right.

Further, these examples illustrate a feature of my account: It can sometimes be hard to know, even for the agent themselves, whether an action has moral worth. This is because it can be hard to know whether your action is explanatorily connected to the actual moral facts. But that moral worth isn't always easy to judge is not a problem, I take it, for an account of moral worth.

There are clearly other possibilities for this connection between rightness and action. But our investigation to coincidence suggests that there needs to be some connection for an agent to act non-coincidentally rightly.

## 7 Some Cases

In this section I'll argue that Unified Explanation RMF deals well with some commonly discussed cases – both cases normally taken to favor the RMF view and those taken to favor the RI view. In particular, it elegantly and plausibly distinguishes between variants of the cases.

### 7.1 Huck Finn

The most discussed case in the moral worth literature is that of Huckleberry Finn – the character from Mark Twain's novel. Huck is a white teenager living in south of the USA in the mid-19th century. He befriends an escaped slave, Jim. At a key point he is conflicted about whether to turn Jim in or to help him escape. He ends up helping Jim escape even though he thinks it is morally wrong since it amounts to stealing from Jim's 'rightful owner'.

The case is a core motivation for the RMF view. It's clear, many think, that Huck helping Jim escape is worthy even though he is not motivated by the thought that it is right. What matters is that Huck is motivated by the right-making features of his action – Jim's humanity and the value of his life.

But, as I've argued, it's not enough that Huck is motivated by right-making features, there needs to be an explanatory connection between the facts about rightness and Huck's action. There needs to be a unified explanation of this matching. And in a natural interpretation of the case there is such an explanation. Huck, it seems, doesn't value Jim's humanity at random. It is not, we want to say, merely coincidental that Huck's valuing Jim's humanity lines up with what really is of value. Rather, Huck has some moral insight that explains his motivations and his ultimate choice to help Jim.

There are, of course, versions of the case where Huck doesn't have such a connection to the moral facts. Sliwa [2016, section 8], for example, describes a version of the case where Huck doesn't turn Jim in because of their friendship. But, of course, the morally relevant fact isn't that Jim is Huck's friend. In this case Huck doesn't seem to have any connection to the moral truth and it does seem accidental that Huck did the right thing – the Unified Explanation RMF gets that result.

But some versions of the case involve a connection to the moral facts. Here, for example, is part of Arpaly's description of the case:

> during the time he spends with Jim, Huckleberry undergoes a perceptual shift... Talking to Jim about his hopes and fears and interacting with him extensively, Huckleberry constantly perceives data (never deliberated upon) that amount to the message that Jim is a person, just like him. Twain makes it very easy for Huckleberry to perceive the similarity between himself and Jim: the two are equally ignorant, share the same language and superstitions, and all in all it does not take the genius of John Stuart Mill to see that there is no particular reason to think of one of them as inferior to the other. While Huckleberry never reflects on these facts, they do prompt him to act toward Jim, more and more, in the same way he would have acted toward any other friend. That Huckleberry begins to perceive Jim as a fellow human being becomes clear when Huckleberry finds himself, to his surprise, apologizing to Jim – an action unthinkable in a society that treats black men as something less than human. As mentioned above, Huckleberry is not capable of bringing to consciousness his nonconscious awareness and making an inference along the lines of 'Jim acts in all ways like a human being, therefore there is no reason to treat him as inferior, and thus what all the adults in my life think about blacks is wrong.' (pp. 76-77)

The natural way to read this, at least for me, is that during his time with Jim, Huck gains a kind of access or connection to moral facts. Arpaly suggests that the fact that 'there is no particular reason to think of one of them as inferior to the other' 'prompts' Huck to act in certain ways towards Jim. And, she seems to say, Huck has a 'nonconscious awareness' of something like the fact that 'Jim acts in all ways like a human being, therefore there is no reason to treat him as inferior, and thus what all the adults in my life think about blacks is wrong.'

The picture, it seems, is that Huck gains some loose, inchoate, access to the moral fact that black people aren't inferior and so should be treated equally. And this is (at least part of) what explains his actions. At least, this is the picture I get from Arpaly's passage that convinces me that Huck's action has moral worth.

There's another reading of the above passage, though, where Huck has no connection to the moral facts but rather gains access to purely descriptive facts. On this reading his 'nonconscious awareness' is of merely descriptive facts about the similarity between him and Jim, or black people more generally, in various descriptive respects. This is, in fact, the reading that fits with Arpaly's official view. Her official view does not require that Huck has any connection to the moral facts in order for his act to be worthy. However, she is at her most convincing in arguing that Huck's action has moral worth when you get the sense that Huck has been able, somehow, to reach out and touch the moral realm.

My account says that on the interpretation where Huck has some kind of unconscious connection to the moral facts, his action is worthy. This, it seems to me, is the right result.

Notice that my view does not require that Huck has knowledge of, or even believes, the moral truths. I agree with Arpaly that it this not required. But the moral truths do need to play a role. While Huck 'never reflects on these facts', they must 'prompt him' to act as he does.

### 7.2 Jean, again

While the Huck Finn case has been a core motivation for the RMF view, the Jean case (and related cases like Herman's [1981, p. 364-5] art thief) have been taken to be a problem for the RMF approach.

The challenge, as we discussed in section 4.1, is distinguishing between variants of the case where Jean acts with worth and where she does not without appealing to counterfactual differences. The Unified Explanation RMF does this by appeal to explanatory differences.

For Jean's giving her friend a ride to work to be worthy there must be a unified explanation of why Jean did the right thing. Perhaps Jean had some moral insight or moral education that explains why she did the right thing. If so we can do more than simply explain why Jean gave her friend a ride and why it was right for her to do so – there is a relevant explanatory connection between rightness and the action.

Notice that if there is such a unified explanation of why Jean did the right thing then, typically, there will be reason to expect Jean to act rightly in related situations – for example, not murdering her friend's ex-boyfriend to prevent embarrassment. But this won't always be the case. Perhaps there is something about Jean's character that means that her moral insight or education doesn't properly transmit to action when it comes to killing ex-boyfriends. So there are some cases, my view suggests, where Jean's act of giving her friend a ride is worthy even if she would murder the ex-boyfriend. There is a sense in which this version of Jean is lucky that she didn't encounter a case where killing an ex-boyfriend was at issue. But this type of luck is, on my view, consistent with her actual action having moral worth.

This is the right result – the fact that there are some possible situations where an agent would do terrible things doesn't undermine the worth of their actions in general. There are people who actually do terrible things and can still act worthily, even if, of course, their character is flawed.

Jean's actions are not worthy in the cases where her giving a friend a ride is not explanatorily connected to the truth. Most cases where Jean is disposed to do bad things in related circumstances – like kill the ex-boyfriend – are like this.

### 7.3 Venom

My RMF view draws plausible distinctions between versions of the **Venom** case.

Jack making the incision in **Venom** doesn't have moral worth. But, Singh discusses a variant, **Venom***, where Jack makes the incision because of his intrinsic desire to save lives, not his intrinsic desire to extract venom. As Singh notes, it seems much more intuitive to ascribe Jack's action moral worth in this case. Similarly, if Jack acts out of an intrinsic desire to improve the welfare of other people his action seems perfectly worthy.

Nevertheless, Singh says that those actions are not worthy. But my view avoids this result. Notice that when Jack is motivated by a desire to help others then it's much more plausible that Jack's action is explained by the moral facts. While an intrinsic motivation to help others is plausibly explained by access to the moral facts, that's not the case with an intrinsic desire to extract venom.

In fact, there is a class of cases like this. Sometimes an agent is motivated by a right-making feature, but that particular motivation gives us evidence that their action is not explanatorily connected to rightness. Our intuitive reaction to such cases are in line with the Unified Explanation RMF.

* * *

My RMF approach handles cases that have traditionally been taken to favor the RMF view – like **Huck Finn** – and those that have been taken to tell against it – like **Jean** and **Venom**. In both cases the appeal to explanatory considerations can draw the line between versions of the case where agents act with worth and where they do not.

## 8 Other RMF accounts

How does the Unified Explanation RMF relate to other RMF views in the literature? It turns out that many RMF views fail to rule out coincidentally right actions. Others do meet the non-coincidentality condition but don't get the results that motivated the RMF view in the first place. Seeing how other RMF views have difficulty finding their way between these pitfalls provides more support for my view. Of course, a full survey of RMF views is not possible, in this section I'll consider a hopefully representative sample.

### 8.1 Arpaly and Markovits

The influential RMF views of Arpaly [2002] and Markovits [2010] both, on the face of it, look to be versions of the Correlational RMF. So they face the problems discussed in section 4, like the **Hiring** case.[14] Markovits's [2010, p. 205] view is that 'my action is morally worthy if and only if my motivating reasons for acting coincide with the reasons morally justifying the action'. No connection between the moral facts and the action is required, merely a certain matching. Arpaly does allow modal considerations to be relevant for moral worth but only for how worthy an action is, not whether the action is worthy. Her account of whether is action is worthy is that 'For an agent to be morally praiseworthy for doing the right thing is for her to have done the right thing for the relevant moral reasons' (p. 84).

[14] Though whether **Hiring** is a problem case for Markovits in particular is complicated by her focus on *subjective* reasons. A full discussion is not possible here but Markovits's (section 3) discussion of the parent who wrongly believes her child swallowed a marble seems to imply that in a version of **Hiring** where Stephanie – the hiring manager who gives the right person the job based on incorrect racial stereotypes – does not normally believe racial sterotypes and typically is a good judge of programming ability she does act with moral worth. But I take this to be the wrong result.

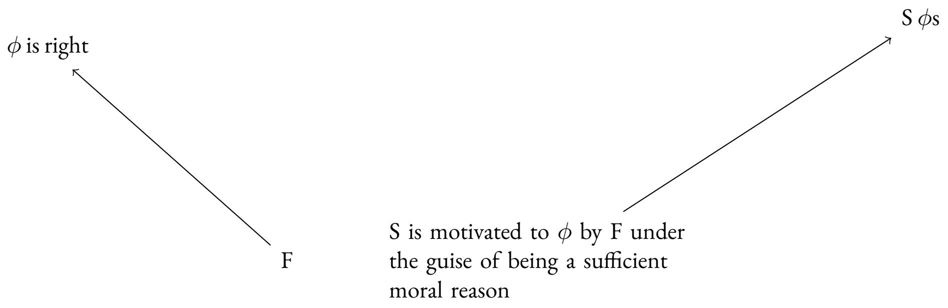
$\phi$ is right
S $\phi$s
F
S is motivated to $\phi$ by F under the guise of being a sufficient moral reason

Figure 7: The Guise of Moral Reasons view

Though there are readings of both Arpaly and Markovits where they understand terms like 'motivating reason' as requiring an explanatory connection between action and the reason. Consequently, they would understand slogans like 'Morally worthy actions are motivated by the features of the action that make the action right' as requiring an explanatory connection between the action and the right-making features. On this reading Arpaly and Markovits would hold versions of the Third-factor RMF view, and so would face the problems with that view.

## 8.2 Singh

Singh [2020, p.161] criticizes Markovits for allowing action that are coincidentally right, appealing to his **Venom** case. But his positive view also allows for coincidentally right actions.

Singh's view is that 'A right action has moral worth if and only if the agent performs it on the basis of sufficient moral reasons as such.' To be motivated by sufficient moral reasons as such is to act under the 'guise' of those reasons being sufficient moral reasons. And 'to act for some reason under the guise of a moral reason is to be motivated by that reason in virtue of taking it to contribute to the overall moral status of the action' (p.170-172). So, it's not enough to act for the right reasons, you need to (in some sense) take those reasons to be the right reasons.

The problem is that this is also a type of correlational view. This is made clear by Singh's claim that the view avoids cases of accidentally doing the right thing 'because in requiring that the agent be motivated by sufficient moral reasons under that very guise, it requires the agent's motivational structure to more closely mirror normative reality' (p. 173). Singh repeatedly talks of 'mirroring' – illustrating how the view postulates a matching between the structure that leads to an agents action, and the structure that leads to an action being right, but it does not postulate any connection between those two sides. See figure 7.

Consider, for example, a slight variant of a case discussed in section 3.1: A person refrains from performing a particular action because it would lead to the killing of another person and this would be wrong. The agent therefore acts upon sufficient moral reasons as such – that the action would lead to killing is sufficient for it being wrong. But if the agent only does this because of their deeply mistaken moral theory – where they value suffering and so think that people should be kept alive longer so they have more opportunities to suffer — then their action isn't worthy. Singh's view doesn't seem to get this result.

But regardless of the details of such a case, that Singh's view faces a problem with coincidence can be seen just by noticing that it is a correlational view.

Singh's view could be adapted to add a connection between the two sides of the coincidence. The natural way to do it, I think, is to add that S being motivated to $\phi$ by F under the guise of being a sufficient moral reason is explained by the fact that F is a sufficient reason to $\phi$. That is, S's motivation is explanatorily connected to, and doesn't just mirror, the facts in the moral domain.

While the resulting view would avoid coincidentality worries it faces another concern with Singh's view – that it gets the wrong results in the Huck Finn case. Huck, it seems, is not motivated to act in virtue of taking some reason 'to contribute to the overall moral status of the action'. That an overly intellectualized picture of the situation. Singh's response is that Huck 'tacitly' takes Jim's personhood to 'constitute sufficient moral reason to help him'. This, I think, is somewhat obscure. A cleaner approach, I think, is mine – where Huck's action is worthy not because he represents certain reasons, tacitly or not, as sufficient moral reasons, but rather because his action is appropriately explained by the moral facts.

### 8.3 Isserow

Isserow [2019] has a disjunctive view where either concern for rightness or right-making features can make for moral worth. For our current purposes we can focus on the right-making features disjunct. This disjunct says that 'a right action is non-accidental in the sense that is relevant to determining its moral worth' if 'the agent acts from a non-instrumental concern for its... right-making features and it is something that she competently brings about... qua action with right-making features' (p. 261).

Exactly what it is to 'competently bring about' something is not obvious. But Isserow, in the spirit of Bradford [2015], claims that 'what is needed is (relevant) justified beliefs; that is, the agent must have justified beliefs about the action that she is performing' (p. 262). In particular, she claims, agents that are motivated by right-making features 'must plausibly be justified in believing that they ought to act as they do'.

This condition is a little strong though – Huck Finn doesn't believe that he ought to act as he does. And we might doubt whether he even has justification for such a belief, regardless of whether he holds it. Even if Huck's actions are explained by an inchoate access to the moral facts, it's not clear that he has justification for believing that he shouldn't turn in Jim, given the strong testimonial evidence he is getting from society as a whole. Though these questions about moral justification are too far afield to get into further.

The larger issue with Isserow's approach is that agents can be justified in believing that their action has certain right-making features, without that explaining why the agent acted.

Consider a variant of **Venom** where Jack, the surgeon, is justified in believing that extracting the venom from the snake-bitten hiker is the right thing to do but he doesn't care about this at all. Rather he extracts the venom because he is simply intrinsically interested in draining venom out of wounds. In fact, we could even assume that Jack is dissuaded from performing the action by his recognition that it's right, but his intrinsic desire to drain venom is so strong that he does it anyway.[15] It's just a coincidence that Jack did the right thing. It's a coincidence that his

[15] Perhaps you think that Jack can't believe that extracting the venom is right without being somewhat motivated to do it. Even so, this doesn't guarantee that the explanation for why Jack extracts the venom is the fact that it was right. For example, it could be that although Jack was somewhat motivated, he wouldn't have extracted the venom for purely moral reasons, because the patient was an enemy of his. However, he does extract the venom because of his strong intrinsic desire to drain venom.

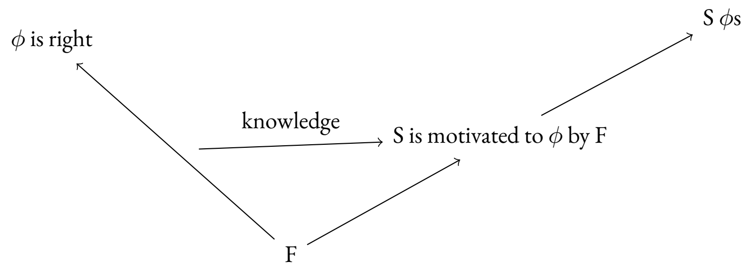

Figure 8: Matching Reasons$_{\text{Kantian}}$

intrinsic desires lined up with what is right.

A possible response is that Jack, in this case, doesn't 'competently bring about' the extraction of the venom qua action with right-making features. Perhaps this is reasonable, but a much richer notion of competently bringing about would be required to cash this out. As it stands, it's unclear what this would look like.[16]

### 8.4 Way

Way's [2017] Matching Reasons$_{\text{Kantian}}$ suggestion does seem to meet the anti-coincidentality condition. It has a similar structure to Isserow's view but involves an explanatory connection between the agent's epistemic state and their action.

> Matching Reasons$_{\text{Kantian}}$: When $\phi$-ing for reason r, you are creditworthy for $\phi$-ing iff (i) r is a reason to $\phi$ and (ii) you $\phi$ because you know that r is a reason to $\phi$.

Condition (ii) postulates an explanatory connection between your knowledge of the moral facts and your action. Assuming you knowing the moral facts implies that your belief is explanatorily connected to the actual moral facts then this establishes an explanatory connection between the moral facts and your action. As such we have an explanation of why you did the right thing that is not merely stapling together distinct explanations. Figure 8 illustrates this structure.

This view avoids coincidentality but at the cost of implying that Huck Finn's actions are not worthy, since Huck doesn't know, or even believe, that any particular reason r is a reason to help Jim.

Way, however, rejects Matching Reasons$_{\text{Kantian}}$.The view he endorses doesn't provide a unified explanation of the why the agent did the right thing. Here is the view:

> Matching Principles: Your $\phi$-ing for reason r is creditworthy iff (i) r is a reason to $\phi$ and (ii) the principle from which you $\phi$ matches a principle which explains why r is a reason to $\phi$.

This view is very closely related to the third-factor RMF view, and, as such allows for the possibility that your reasons for acting merely mirror the reasons why the action is right without there being a unified explanation. See figure 9.

[16] As we will see later, one promising approach is to give an interpretation of 'competently bring about' an action qua action with right-making features that implies my account. As such you could fit my view into Isserow's framework.

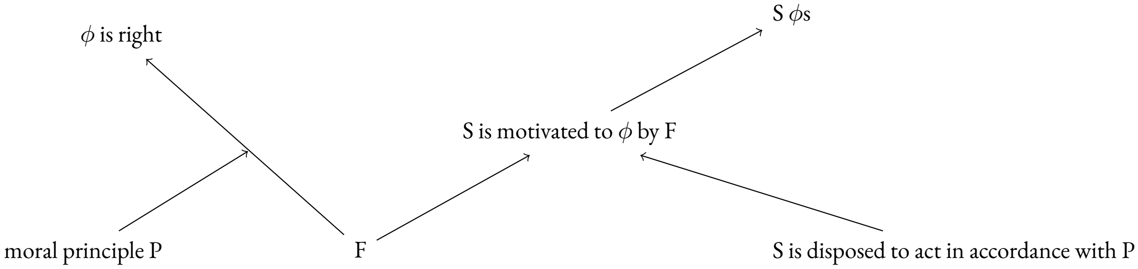
S $\phi$s
$\phi$ is right
S is motivated to $\phi$ by F
moral principle P
F
S is disposed to act in accordance with P

Figure 9: Matching Principles

### 8.5 Know-How

Some recent approaches to moral worth, for example, Lord [2017] and Cunningham [2021] are based on the idea of *know-how*. The rough idea is that an action is worthy if it is motivated by the right-making features and it's a manifestation of know-how about how to respond to those features. Of course, what it is to manifest know-how is not obvious.

Glossing over some nuances, the core of Cunningham's view is that to act with moral worth you have to be motivated by some right making feature F that you have knowledge of and your action has to manifest know-how about how to respond to reasons of type T, where F is type T.

Cunningham suggests that part of manifesting the relevant know-how involves understanding normative truths. For example, when discussing a case he says 'Her misunderstanding of the normative significance of promise-keeping is indeed so bad, we should deny that she knows how to respond to such considerations'. The problem is that this excludes most versions of the Huck Finn case. Cunningham, of course, recognizes this and tries to defend the idea that most versions of Huck do not act with worth. I'm not convinced, but that's too much to get into here. (A further concern with Cunningham's approach is that the appeal to how the agent would respond to other reasons of type T is plausibly a violation of the pertinence constraint.)

Lord's view is in a similar spirit. But, notably, he suggests that when agents act with worth there is an explanatory connection between between moral facts and action. He claims that manifesting know-how involves dispositions to act that are sensitive to the actual moral facts. For example, Immanuel the grocery store owner should be disposed 'to give $1.00 back [to the customer] when the fact that $1.00 is the correct change provides a sufficient (moral) reason to give $1.00 back' (p. 458). Further, he suggests, it's 'plausible' that when there is such a disposition Immanuel's action of giving the customer $1.00 is 'caused' by the fact that $1.00 being the correct change is a sufficient reason to give it to the customer. So, Lord claims that it's plausible that in such cases there is a causal relationship between the agent's action and the moral facts. Presumably this implies an explanatory connection too.

This isn't the whole story about manifesting the relevant know-how[17] in addition it has to be the case that the agent is in a position to know that the relevant feature F is a sufficient reason to perform the action. So, Immanuel has to be in a position to know that the fact that $1.00 is the correct change is a sufficient reason to give it to the customer.

This condition seems to rule out Huck Finn from acting with worth. He's not, in seems, in a position to know that

[17] In fact, Lord doesn't claim to give a full story about manifesting know-how.

he should help Jim and not turn him in. His location within his society puts him in a very poor epistemic position with respect to the relevant moral facts. Only, it seems, with a very significant change in his epistemic position could he know them. (See Lord [2018, section 3.6] for his discussion of 'position to know'.) But the issues here are complicated and Lord [2017] expresses openness to other epistemic conditions replacing 'being in a position to know'.

Of the existing views I think Lord's view does the best job at meeting the anti-coincidentality condition in a way that might be consistent with Huck Finn's actions having moral worth. Lord's view is plausibly understood as embodying an explanatory conception of non-coincidentality. Lord clearly intends to appeal to a conception of disposition which is weighty enough to generate an explanatory connection between the normative facts and agents' actions (see Lord [2018, section 5.4 and 5.5.3]). And this explanatory connection goes beyond merely stapling together explanations of why the agent did $\phi$ with an explanation of why $\phi$ is right.

Further, Lord's view doesn't build in a highly intellectualized picture of what must be going on in an agent's head for their actions to have worth so perhaps it can get appropriate results about Huck Finn.

The major concern with Lord's view is somewhat different (and applies equally to Cunningham's view). The appeal to dispositions violates, at least the spirit of, the pertinence constraint. In particular, it seems that his view rules out moral worth for 'out of character' actions, where the agent faces a kind of case where they are not generally disposed to act rightly but yet do so. As cases like Markovits's fanatical dog lover suggest, if an agent acts rightly and is appropriately connected to the moral facts then the fact they would act badly in related situations relects badly on their character, but not on the worth of this action. Similarly, if a beginner new chess is generally disposed to calculate poorly, but, in this case, they calculate a line correctly and so play the best move then that move is creditworthy even if them playing the right move was very modally fragile.[18]

But again, Lord's view is successful in ruling out coincidentally right actions while finding space for Huck Finn's actions to be worthy.

* * *

RMF views face problems with avoiding coincidence while retaining the spirit of the account. Strategies to to avoid coincidentally right actions often threaten to undermine key motivation for the RMF account in the first place, by implying that (most versions of) Huck Finn do not act with moral worth. The RMF approach I suggested, based on understanding what is required for an explanatory conception of non-coincidentality, finds a way between these pitfalls.

## 9 Deviant Explanatory Chains

I have been developing an RMF account based on the idea of avoiding coincidence. The Unified Explanation RMF gives a necessary condition for moral worth. But what is needed for sufficiency?

[18] A possible response is that the relevant dispositions can be extremely fragile. But, if we strip dispositions of what seems to be their core feature – their connection to modality – then it becomes hard to get a grip on what the appeal to dispositions is supposed to do for us.

Remember that our initial intuition was that actions need to be non-accidentally right to be worthy. Non-accidentality entails non-coincidentality, so we can get a necessary condition on moral worth by understanding coincidence. But, plausibly, there are cases where the agent acts non-coincidentally rightly – since there is a unified explanation of why the agent did the right thing – but still accidentally rightly. The threatening cases are where the explanatory connection between rightness and action is of a deviant kind.

What might these cases look like? Let's consider a few possibilities. Start by considering an epistemic, not a moral, case.

> **Fake Fake Barn Country** You are driving through a region with lots of barn-facades – which from the road look exactly like barns. However, unbeknownst to you, hidden just behind every one of these barn-facades is a real barn that cannot be seen from the road. You look towards one of these barn-facades and form the belief 'there's a barn over there'.

There is a matching between your belief and the truth – there really is a barn over there. And there is a unified explanation of why you believed correctly – it's no coincidence that your belief was true. But, still, there is a sense in which you only accidentally believe the truth. The explanatory chain between truth and belief is deviant. Intuitively, your belief that there's a barn over there does not constitute knowledge.[19]

However, I don't think such cases are problems for my view. My judgement is that your belief that there's a barn over there is epistemically worthy.[20][21] There might be a sense in which the agent is only accidentally correct, but I don't think this is the relevant sense of accidentality for moral worth. Cases of this structure do not identify the gap between coincidence and the relevant sense of accident.

Here's another possible case of a non-coincidentally but accidentally right action. Take a politician who does charity work which improves their electoral chances. She is motivated by the fact that some people are in need and that she is in a position to help. But she only has this motivation because of the way it helps her electoral chances – the public are more likely to vote for a politician who acts morally. So, it's non-coincidental that she did the right thing, there is a unified explanation that goes via her desire to be elected and the public's desire for moral politicians. But, her action is, in the relevant sense, accidentally right.

Such cases are easily dealt with by RMF accounts – we just need to clarify that agents should be *non-instrumentally* motivated by the right-making features (see e.g. Markovits [2010, p.230]. Isserow [2019, section 2]). So we should add adapt condition (i) of Unified Explanation RMF correspondingly. But this case doesn't raise any particular worry about the gap between accident and coincidence.

Here's a third possible non-coincidentally but accidentally right action – a Frankfurt-style case. Consider Jack in **Venom**, who does the right thing for a right-making reason but only accidentally does the right thing. Then add that there is some demon who can directly intervene with Jack's brain and would make Jack do the right thing if he was about to act wrongly. As it happens, though, Jack does the right thing and the demon does not intervene. Jack

[19] Cunningham's [2021] *Masochist II* case is a moral example with a similar structure.

[20] Though I don't commit to the claim that the belief constitutes knowledge.

[21] I make the analogous judgment in the *Masochist II* case, contra Cunningham.

accidentally does the right thing, in the sense of 'accidental' relevant to moral worth. But, we might worry, it's not a coincidence that Jack does the right thing, since there demon provides a robust explanation of why he will always do the right thing.

However, this isn't, in fact, a case of Jack non-coincidentally doing the right thing. The fact that Jack will always do the right thing – because of the demon – does not establish that he does the right thing non-coincidentally. What matters is the actual explanation of the matching between Jack's action and rightness. And in the actual case the demon is not explanatorily relevant so Jack coincidentally does the right thing, just as in the normal **Venom** case.

In general, we don't take backups to be part of the actual explanation of an event. Imagine that Suzy throws a rock at a window, breaking it. If she hadn't thrown then Billy would have thrown the rock. Suzy's action explains the window breaking – Billy is a mere backup. Similarly, the demon is a mere backup.[22]

In principle, it looks like the gap between coincidentality and accidentality leaves cases that are not covered by Unified Explanation RMF. But it's somewhat hard to find such cases. It's hard, that is, to find cases where an agent acts rightly; is non-instrumentally motivated by features which actually make the action right; where there is a unified explanation of the matching between rightness and action; and where the agent's action clearly doesn't have moral worth.[23]

But here is a case that seems to meet these conditions: Imagine an agent recognizes that some particular action has right-making features and becomes motivated to do the complete opposite because they are an anti-moralist. Then a demon interferes with their brain and flips their motivation so now they are motivated to perform the action. The agent seems not act with moral worth even though they are motivated by right-making features and there is a unified explanation (via the actions of the demon) of why they acted rightly.

This shows that the Unified Explanation RMF is only a necessary condition on moral worth. But it's hard to know how worried to be about cases like this. These cases invite the question of what we are even trying to do with an account of moral worth. Are we really trying to give a totally complete set of necessary and sufficient conditions that rule out all counterexamples, however strange? Or are we just trying to identify the core of what makes the actions we see around us worthy or not. Understood in this latter way, my account says that the core of moral worth is motivation by the right-making features and the existence of a unified explanation that makes it non-coincidental that the agent acted rightly. That's not *merely* a necessary condition on moral worth – it's an identification of *what matters for moral worth*. My sense is that most of the literature should be understood in this latter way.

Even if we understand the project in the former way, notice that similar cases lead to worries for RI views. Imagine an agent recognizes the rightness of an action and then becomes motivated to do the complete opposite because they are an anti-moralist. Then a demon interferes with their brain and flips their motivation so now they are motivated to perform the action. If our view on moral worth is based upon the motivations that an agent has, as both the RI

[22] Notice that this variant **Venom** case is, in the spirit of Frankfurt cases more generally, a problem for modal conditions on moral worth, not explanatory conditions. The demon could make it such that Jack could not easily have acted wrongly, or could impose other modal conditions on Jack's action. But still, it's clear that Jack's action doesn't have moral worth. Actual world relations, not modal conditions, are important for moral worth.

[23] Remember, one issue that we are putting aside is just how many of the right-making features an agent need to be motivated by to act with worth (see footnote 10).

and RMF views are, then it's is not surprising that a demon fiddling directly with people's motivations can cause problems.

These cases show us that neither the RMF nor the RI view has escaped the problem of deviant explanatory chains. If we want to generate necessary and sufficient conditions, then, we will need to add some stipulation or machinery in order to rule out such deviant chains. Some views in the literature, both RI and RMF do this. For example, the RI view of Johnson King [2020] involves the idea that an agent must *deliberately* do the right thing. Part of the role of deliberateness seems to be to rule out deviant chains. Similarly with Isserow's *competently bringing about* and the concept of *manifesting know-how* that Lord [2017] and Cunningham [2021] stress. But such machinery doesn't really solve the problem of deviant explanatory chains – rather it just labels the problem.[24]

That's not to say that *deliberateness* and *competently bringing about* and so on are black boxes. The respective authors tell us a lot about these concepts. But it's as if these concepts are fairly transparent boxes, with their machinery mostly on view, except the component that deals with deviant chains is hidden.[25]

To be clear, this is not a criticism! It's totally reasonable to label the problem, given that the problem of deviant chains is a deep issue across a variety of debates in philosophy. In fact, labelling the problem is a way to bring together the two approaches to the moral worth project I just mentioned. The central aim, I think, is this second project of identifying the core of moral worth – what makes for moral worth in ordinary cases, and in interesting cases like that of Huck Finn. While that is our focus we can label the gap left by the possibility of deviant explanatory chains.

There are a few ways of implementing this strategy with respect to my account. Perhaps the easiest is to take on Isserow's locution of *competently bringing about*. Thus an agent's action is morally worthy if and only if they are non-instrumentally motivated by right-making features of the action and they competently bring about the action qua action with right-making features. However, we would specify that competently bringing about an action qua action with right-making features implies the non-coincidentality condition that I defended – that is, there has to be a unified explanation of why the agent did the right thing.

But, again, the core of what makes for moral worth, on my view, is given by Unified Explanation RMF. It's just can we can, if we desire, massage that into necessary and sufficient conditions.

## 10 Conclusion

Morally worthy action requires that the agent does the right thing non-accidentally. That implies that the matching between rightness and the action performed is non-coincidental. Understanding the nature of coincidentality allows us to see that rightness itself views of moral worth should be formulated in explanatory terms. Modal or merely correlational versions of the view aren't aren't adequate.

Similarly, right-making feature views should be given an explanatory formulation. But it must be the right kind of explanation. It's not enough to just staple together explanations of why the agent $\phi$-ed and why $\phi$ was right. What

[24] One might argue that these concerns about deviant chains are a reason to reject explanatory conceptions of non-accidentality in favor of modal conceptions. I don't think this is right for a variety of reasons but, most importantly, as we pointed out in footnote 22, similar cases are problems for the modal conditions on non-accidentality.

[25] In Lord's [2018] discussion of deviant chains he notes that 'What I've done is isolated where the problem lies.' (p. 138)

is needed is a *unified* explanation – an explanation of the matching between action and rightness that is better than merely explaining the rightness and the action separately.

Adding this anti-coincidentality condition to a simple RMF view generates a very attractive approach to moral worth – the Unified Explanation RMF. Past RMF views have faced difficulties ruling out coincidence in a way that retains core intuitions – about cases like Huck Finn – that motivated RMF views in the first place. My account finds a way between those problems. There is still the possibility of some strange deviant explanatory chains, but all approaches to moral worth face this. The Unified Explanation RMF identifies the core of what makes actions worthy.

The success of the Unified Explanation RMF – particularly that it avoids accidentality worries as well as RI views – is a powerful argument against RI views. It makes it rather hard to see we would accept an RI view in its place. And, given the way that the Unified Explanation RMF takes on part of the spirit of the RI account, perhaps those who have been attracted to the RI account might embrace this RMF view.

## References

Nomy Arpaly. *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press, November 2002.

Dan Baras. *Calling for Explanation*. Oxford University Press.

Dan Baras. How can necessary facts call for explanation? *Synthese*, pages 1–18, 2020.

Sharon E Berry. Coincidence avoidance and formulating the access problem. *Canadian journal of philosophy*, 50 (6):687–701, August 2020.

Harjit Bhogal. Explanationism versus modalism in debunking (and theory choice). *Mind; a quarterly review of psychology and philosophy*.

Harjit Bhogal. Coincidences and the grain of explanation. *Philosophy and phenomenological research*, 100(3):677–694, 2020.

Harjit Bhogal. What's the coincidence in debunking? *Philosophy and phenomenological research*, July 2022.

Gwen Bradford. *Achievement*. Oxford University Press, 2015.

Joe Cunningham. Moral worth and knowing how to respond to reasons. *Philosophy and phenomenological research*, 105(2):385–405, 2021.

David Faraci. Groundwork for an explanationist account of epistemic coincidence. *Philosophers Imprint*, 2019.

Hartry Field. The a prioricity of logic. In *Proceedings of the Aristotelian Society*, volume 96, pages 359–379, 1996.

Daniel Fogal and Alex Worsnip. What the cluster view can do for you. In *Oxford Studies in Metaethics, Volume 19*.

Herbert Hart and Tony Honoré. *Causation in the Law*. OUP Oxford, May 1985.

Barbara Herman. On the value of acting from the motive of duty. *The Philosophical review*, 90(3):359, July 1981.

Barbara Herman. *The Practice of Moral Judgment*. Harvard University Press, 1993.

Nathan Robert Howard. One desire too many. *Philosophy and phenomenological research*, 102(2):302–317, 2021.

Jessica Isserow. Moral worth and doing the right thing by accident. *Australasian journal of philosophy*, 97(2): 251–264, April 2019.

Jessica Isserow. Doubts about duty as a secondary motive. *Philosophy and phenomenological research*, 105(2):276–298, 2021.

Zoe Johnson King. Accidentally doing the right thing. *Philosophy and phenomenological research*, 100(1):186–206, 2020.

Jaegwon Kim. *Supervenience and Mind*. Supervenience and Mind. Cambridge University Press, Cambridge, January 1993.

Daniel Z Korman and Dustin Locke. Against minimalist responses to moral debunking arguments. *Oxford Studies in Metaethics*, 15, 2020.

Tamar Lando. Coincidence and common cause. *Noûs*, 51(1):132–151, March 2017.

Marc Lange. What are mathematical coincidences (and why does it matter)? *Mind*, 119(474):307–340, July 2010.

Øystein Linnebo. Epistemological challenges to mathematical platonism. *Philosophical studies*, 129(3):545–574, June 2006.

Errol Lord. On the intellectual conditions for responsibility: Acting for the right reasons, conceptualization, and credit. *Philosophy and phenomenological research*, 95(2):436–464, September 2017.

Errol Lord. *The Importance of Being Rational*. Oxford University Press, 2018.

Matt Lutz. What makes evolution a defeater? *Erkenntnis*, 83(6):1105–1126, 2018.

Paolo Mancosu. *Explanation in Mathematics*. Metaphysics Research Lab, Stanford University, summer 2018 edition, 2018.

Julia Markovits. Acting for the right reasons. *The Philosophical review*, 119(2):201–242, 2010.

Robert Nozick. *Philosophical Explanations*. Philosophical Explanations. Belknap Press, Cambridge, MA, January 1981.

David Owens. *Causes and Coincidences*. Cambridge University Press, January 1992.

Jonathan Schaffer. On what grounds what. In David Manley, David J Chalmers, and Ryan Wasserman, editors, *Metametaphysics: New Essays on the Foundations of Ontology*. OUP, January 2009.

Theodore Sider. *The Tools of Metaphysics and the Metaphysics of Science*. Oxford University Press, January 2020.

Keshav Singh. Moral worth, credit, and Non-Accidentality. In Mark Timmons, editor, *Oxford Studies in Normative Ethics, Vol. 10*. Oxford University Press, 2020.

Paulina Sliwa. Moral worth and moral knowledge. *Philosophy and phenomenological research*, 93(2):393–418, September 2016.

Philip Stratton-Lake. *Kant, Duty and Moral Worth*. Routledge, New York, 2000.

Sharon Street. A darwinian dilemma for realist theories of value. *Philosophical studies*, 127(1):109–166, 2006.

Sharon Street. Objectivity and truth: You'd better rethink it. *Oxford studies in metaethics*, 11(1):293–334, 2016.

Jonathan Way. Creditworthiness and matching principles. In Mark Timmons, editor, *Oxford Studies in Normative Ethics, Vol 7*. Oxford University Press, 2017.

Tobias Wilsch. Sophisticated modal primitivism. *Philosophical Issues. A Supplement to Nous*, 27(1):428–448, January 2017.