**Knowing-to in Wang Yangming**

Waldemar Brys

University of New South Wales

Accepted Manuscript

The Version of Record is forthcoming in: *Oxford Handbook of Chinese Philosophy*, edited by Justin Tiwald. New York: Oxford University Press.

**Abstract**

Wang Yangming 王陽明 (1472 – 1529) is famously associated with the view that knowledge and action are unified (*zhī xíng hé yī* 知行合一). Call this the *Unity Thesis*. Given standard assumptions about what it means for a person to know, it may seem that the *Unity Thesis* is clearly false: I can know that *p* without currently acting in *p*-related ways, and I can know how to *φ* without currently *φ*-ing. My aims in this paper are, first, to draw on recent work in epistemology to explain and defend the *Unity Thesis* and, second, to argue that it offers us an attractive conceptual alternative to a standard way of thinking about the nature of intentional action. The first step of my argument draws on the idea that what distinguishes intentional actions from bodily events is the presence of knowing-to – that is, an agent is *φ*-ing intentionally if and only if she is currently doing something because she knows to do it as a way for her to *φ*. Such a notion of knowing-to allows us to explain otherwise puzzling features of the *Unity Thesis*, including the claim that knowing and acting occur simultaneously. In the second step, I argue that the *Unity Thesis* can help us avoid various long-standing issues in the philosophy of action, including the problem of deviant (formal) causation.

**Knowing-to in Wang Yangming**

Waldemar Brys

**1. Introduction**

Wang Yangming 王陽明 (1472 – 1529), also known as Wang Shouren 王守仁, is perhaps best known for his theory that knowledge and action are unified (*zhī xíng hé yī* 知行合一). I will henceforth call this the *Unity Thesis*. My aims in this paper are, first, to draw on recent work in epistemology to explain and defend the *Unity Thesis* and, second, to argue that it offers us an attractive conceptual alternative to a standard way of thinking about the nature of intentional action. Part of the reason why it is attractive is that it helps us avoid various long-standing issues in the philosophy of action, including the problem of deviant (formal) causation.

An important aim of much contemporary philosophy of action is to explain the difference between actions performed intentionally, say, when Farmer Zhang is cutting onions to make dinner for his family, and actions performed unintentionally, say, when a person is serving rice and drops some of it on the floor. A popular suggestion is that Farmer Zhang’s action of cutting onions is intentional, because it appropriately matches up with certain *purely inner* mental states of his, such as his intention to cut onions or his belief that what he is currently doing amounts to cutting onions. These mental states are “purely inner” in the sense that they can be present without any specific action being performed.[[1]](#footnote-1) For example, Farmer Zhang’s intention to cut onions is purely inner, because Farmer Zhang can be intending to cut onions without, in fact, currently doing so (because, say, he cannot find his knife). Hence, the purely inner component of Farmer Zhang’s action has to fit his bodily movements in the right way for his action to count as intentionally performed. The problem with such an analysis of intentional action is that it is notoriously difficult to give a satisfying account of what that “right way” is supposed to be.

I return to this issue in the course of my paper, and I argue that Wang Yangming provides us with a conception of intentional action that avoids the issue of having to specify the “right way” that purely inner mental states match up with overt bodily movements. For Wang, it is *knowledge* that plays a special role in explaining human action, rather than an agent’s intentions, desires, or beliefs. Whereas an intention, desire, or relevant belief is a purely inner mental state – in the sense that having it does not entail that you are *φ*-ing – the *Unity Thesis* implies that you cannot be *φ*-ing without (relevantly) knowing, and you cannot be (relevantly) knowing without *φ*-ing. Therefore, according to Wang Yangming, the relevant kind of knowledge is not “purely inner” in the sense that it can be had without acting.[[2]](#footnote-2) Part of the aim of this paper is to explain what this means and to defend it as a plausible conceptual option.

Wang Yangming’s *Unity Thesis* is as famous as it is controversial, and it is not my aim to settle the debate on how exactly we should interpret it. My focus in this paper lies on examining what lessons we can learn about the role of knowledge in action from Wang’s *Unity Thesis*. However, I take my argument to contribute to the broader debate on Wang Yangming exegesis in at least two ways. First, unlike some commentators, I take the *Unity Thesis* to hold true not only of morally virtuous actions, but of all intentional actions. Hence, I take Wang Yangming to be saying that knowledge is a condition on intentional action.[[3]](#footnote-3)I offer an argument for this in Section 2.

Second, Wang Yangming makes clear that it is not *all* knowledge that is unified with action, but only a specific kind of knowledge. He calls it “genuine knowledge” (*zhēn zhī* 真知; IPL 5, 125, 133, 168, 170, 189), knowledge in its “original condition” (*běn tǐ* 本體; IPL 133, 165; cf. IPL 155), or “extended knowledge” (*zhì zhī* 致知; IPL 138, 139, 140, 321).[[4]](#footnote-4) Hence, for Wang Yangming, only some knowledge is unified with action – and I will henceforth call it *knowledge-in-action*. Wang takes such knowledge-in-action to have some sort of primacy over other kinds of knowledge, although the question of how exactly they are related is beyond the scope of this paper.[[5]](#footnote-5) My aim is to show that Wang’s notion of knowledge-in-action can be helpfully explicated in terms of the concept of *knowing-to*: an agent has knowledge-in-action if and only if she knows to do *this* for her to *φ*, where “this” stands for something that the agent is currently doing.

Such an approach agrees with an influential line of thought defended in a series of papers by Huang Yong (2017, 2021, 2022a, 2022b) – namely, that the *Unity Thesis* concerns a kind of knowledge that is in some ways distinct from knowledge-that, knowledge-how, or knowledge of things, and that we should label it “knowing-to.” But I have doubts whether Huang’s notion of “knowing-to” is as helpful as it may seem, and I give reasons for preferring a different conception of knowing-to in Section 3.

Finally, in Section 4, I argue that the *Unity Thesis* offers us a promising alternative to a standard picture on which actions are distinguished from bodily events by the mental states that cause them.

**2. Which Actions Are Unified with Knowledge?**

Part of the aim of this paper is to argue that the following view is plausible. For Wang Yangming, if an action is an intentional action, then it is unified with knowledge. I therefore take it to be plausible that Wang’s *Unity Thesis* entails the view that knowing is necessary for intentionally acting.

Against this, one might object that Wang restricts the scope of the *Unity Thesis* to only apply to those actions that are morally virtuous, rather than to all intentional actions.[[6]](#footnote-6) But I do not think this is right. I agree that Wang takes all virtuous actions to be actions that are unified with knowledge, but I disagree that he *only* takes virtuous actions to be unified with knowledge.My evidence for this comes from one of Wang’s arguments for his *Unity Thesis*:

No one really learns anything without carrying it into action. Take the learning of filial piety. One must relieve one’s parents of the burden of toil, serve and care for them, and personally put the principle of filial piety into action before one can be said to be learning filial piety. Can merely talking about it in a vacuum be considered as learning? To learn archery, one must hold out the bow, fix the arrow to the string, draw the bow, and take aim. To learn writing, one must lay out the paper, take the brush, hold the inkwell, and dip the brush into the ink. In all the world, nothing can be considered learning that does not involve action. […] Therefore if we realize that no learning can be considered learning if it is not carried into action, […] we know that knowledge and action are a unity and advance simultaneously, and cannot be separated. (IPL 136)

In this passage, Wang argues that knowledge and action form a unity, because a person’s knowledge-in-action is improved at the same time as her acting is improved – that is, the knowledge-in-action that features in a person’s *φ*-ingimproves the better she gets at *φ*-ing.[[7]](#footnote-7) What is relevant for my argument is that Wang claims that this applies to knowing *archery* as much as it applies to knowing filial piety. If Wang takes the simultaneous improvement of knowing and acting to support the *Unity Thesis*, then the argument from IPL 136 suggests that some non-moral acting forms a unity with knowing in the same way that morally virtuous acting does. Therefore, it is plausible that Wang takes the *Unity Thesis* to hold true of some non-moral actions as much as it holds true of morally virtuous actions.

Although it is correct that morally virtuous actions frequently appear in IPL as examples of the kind of actions that are relevantly unified with knowledge (e.g., IPL 5, 134, 140, 291), it would be too quick to conclude that Wang takes the *Unity Thesis* to *only* apply to morally virtuous actions.

A different objection one might raise against my proposal is that it is not obvious that Wang takes *all* intentional actions to involve knowing-in-action. Although Wang’s argument for the *Unity Thesis* at IPL 136 involves an appeal to intentional actions, i.e., the act of deliberate practice involved in becoming a better archer, one might wonder if, for Wang, some intentional actions are not unified with knowledge.

It is more difficult to prove that Wang took all intentional actions to be unified with knowledge-in-action, because Wang’s focus is not to give us a theory of intentional action. But I believe that there is some room for optimism here. At IPL 5, Wang makes the following observation: “knowing is acting’s intent (*xíng de zhǔ yì* 行的主意) and acting is knowing’s practice”.[[8]](#footnote-8) If all actions relevant to the *Unity Thesis* are actions with “intent”, then it seems to me at least plausible that they are intentional actions. Hence, all intentional actions are unified with knowledge-in-action, and having knowledge-in-action is a necessary condition for acting intentionally.

The challenge is to make sense of Wang’s notion of knowledge-in-action. What is it that a person supposedly knows when she is performing an action that is unified with her knowledge? I turn to this issue in the next section.

**3. Knowing-in-Action and Knowing-to**

In the first section, I have pointed out that Wang’s *Unity Thesis* is not a thesis about *all* knowledge, but only about what he calls “genuine” or “extended knowledge” – namely, knowledge-in-action. In this section, I argue that the relevant kind of knowledge is best understood as *knowing-to*: S’s *φ*-ing is unified with her knowingif and only if S is knowing to do *this* for her to *φ*.[[9]](#footnote-9)

Knowing-to differs conceptually from knowing-that and knowing-how in ways that I discuss further below. Moreover, knowing-to also differs from the kind of knowledge that Huang Yong has recently proposed as a candidate for interpreting the *Unity Thesis*. According to Huang (2017, 2021, 2022a, 2022b), only knowledge that necessarily motivates its possessor to act is unified with the agent’s action. Huang argues that all such knowledge involves propositions with an “ought”-clause: if I possess knowledge unified with action, then I know that I ought to *φ* and I am thereby motivated to *φ*.[[10]](#footnote-10) I return to Huang’s view further below. For now, I offer an argument for why I believe that Wang’s *Unity Thesis* is best understood in terms of knowing-to. My argument is that knowing-to exhibits many of the characteristics that Wang attributes to the kind of knowledge that is unified with action.

**3.1 No knowing without acting**

A central implication of the *Unity Thesis* is that you cannot be knowing-in-action without acting:

People today separate knowing and acting into two distinct tasks and think that one must first know and only then can one act. They say, ‘Now I will perform the task of knowing, by studying and learning. Once I have attained real knowledge, I then will pursue the task of acting.’ And so, till the end of their days, they never act, and till the end of their days, they never know. [...] My current teaching regarding the unity of knowing and acting is a medicine directed precisely at this disease. (IPL 5)

Wang is clear that the *Unity Thesis* is directed against the view, advocated by some of his contemporaries, that knowing-in-action can temporally precede acting. In response, he writes elsewhere at IPL 5 that, “knowledge is the beginning of action” (*xíng zhī shǐ* 行之始) and “action is the completion of knowledge” (*zhī zhī chéng* 知之成), so that “[t]here never have been people who are knowing but do not act”.[[11]](#footnote-11) If a person’s relevant knowing is the beginning of her corresponding action, then I take it that her relevant knowing is sufficient for acting. Hence, an instance of knowing-in-action cannot temporally precede a corresponding instance of acting.

But it is clear that knowledge-that and knowledge-how *can* temporally precede at least some of the actions that manifest them. I can know that *p* without currently acting in any *p*-related ways. For example, I can know that Sparta won the Peloponnesian War even if I have never told anyone about it. Likewise, I can know how to cook an omelette without currently making one. Commentators on Wang Yangming have struggled to make sense of his claim that possessing relevant knowledge is sufficient for acting, and I believe that part of the reason might be the widely held and plausible assumption that knowledge is a standing state that a person can be in without having to exercise it. It is not that you *cease* to know that *p* when you are not currently performing *p*-related actions.

Knowing-to, however, is only present when the relevant action is. Adapting one of Wang’s examples from IPL 136, suppose you have fixed an arrow on a bow string and are drawing the bow with the aim of hitting a target in front of you. As you are drawing the bow, you might know to draw the bow *this* far for you to hit the target, given the specific details of your situation – for example, the distance of the target away from you, the presence and strength of the wind, the size of the bow relative to your height, etc. The claim is that, if you are knowing to do *this* (i.e., draw the bow *this* far), then you are knowing-to. If you are not drawing the bow far enough for you to hit your target, or if you are merely guessing how far to draw the bow without knowing it, then you are not *knowing to* draw the bow far enough.

Given that what it means to draw the bow *this* far to hit a target is situation-specific, it is plausible that the content of knowing-to is only available to the agent as she is performing the relevant action.[[12]](#footnote-12) If an archer must know to draw the bow *this* far to hit her target under *these* specific circumstances, then she will not have the required situation-specific knowledge when she is currently at home reading a book. It is not that the archer carries in her head a maximally fine-grained blueprint of how far to draw a bow given the specific details of various situations. Put simply, an agent’s knowing-to does not temporally precede her corresponding action, and hence it offers us an elegant way of making sense of an important aspect of Wang’s *Unity Thesis*.[[13]](#footnote-13)

Moreover, such a proposal can also help explain why Wang says that knowledge-in-action is the *beginning* of a corresponding act. If a person is in apt circumstances for her to *φ* and she knows how to *φ*, then that is not enough for her to *φ*, because she must also put her know-how into practice. To put know-how into practice, what is required is knowing-to (Hetherington 2022a: 18-20) – more specifically, without your knowing to do *this* to *φ*, where “this” involves an enactment of a way for you to *φ*, it is hard to see how your know-how can be applied in your *φ*-ing. The moment that you know to cut onions to make dinner on a particular occasion, you have begun making dinner, because you have begun enacting part of a way for you to make dinner, and hence you have begun putting relevant know-how into practice. Therefore, it is plausible to conceive of knowing-to as the beginning of an agent’s *φ*-ing, whereas it is implausible for knowing-that or knowing-how to be so conceived.[[14]](#footnote-14)

One might object that there is a way to take the expression “S is knowing to *φ*” that is incompatible with my argument. For example, one might claim that it is perfectly intelligible to say: Farmer Zhang knew to be polite to the government’s envoy, but he did not know what it meant to be polite in that particular situation, and so he was not polite when greeting the envoy. In this case, Farmer Zhang did not *know* *how* to be polite and so he failed to be polite, even though he knew *to* be polite. Hence, one might object that some instances of “S knows to *φ*” are instances of S’s knowing that she *ought to* *φ*.

That may very well be so. My proposal does not rule out that there is a way of taking “knowing to” in terms of the agent’s knowing that she ought to do something. The suggestion is rather that not all cases of “knowing to” are a species of knowing-that. Therefore, although Farmer Zhang knows that he must be polite to the envoy, and he knows how to be polite towards envoys, it turns out that, in this instance, as Farmer Zhang was greeting the envoy, he did not *know to* greet the envoy in *this* way for him to greet politely.

In response, could one suggest that the reason why Farmer Zhang does not know to greet the government’s envoy in a polite way is because he does not know that he *ought* to greet the envoy in *this* way? No, because that gets the modality of knowing-to wrong. For example, there are many ways for you to dice an onion. If you know to perform *these* specific bodily movements to dice an onion, then it does not follow that moving the knife at this angle, holding the onion with this specific grip, cutting at this speed, etc., are actions that you *ought* to perform to dice the onion. Knowing to do *this* to dice an onion entails knowing to dice it in one (but not the *only* one) of the available ways of doing so. Hence, knowing to do *this* for you to dice an onion does not entail knowing that you *ought* to do *this* for you to dice an onion, because it is simply false that you *ought* to move the knife at this specific angle for you to cut the onion. In a similar vein, knowing that you ought to do *this* for you to *φ* does not entail knowing-to. I can know that I ought to fry eggs to make an omelette without knowing how to fry eggs, and hence without knowing to do *this* (e.g., heat up the pan, melt the butter) to make the omelette.

**3.2 Acting and knowing occur simultaneously**

Wang Yangming offers the following argument in support of his *Unity Thesis*:

Knowing pain offers another good example. One must already be in pain oneself in order to know pain. [Similarly,] one must already be cold oneself in order to know cold, and one must already be hungry oneself in order to know hunger. How can knowing and acting be separated? (IPL 5)

The argument seems to be this. The relation between a person’s (relevant) knowing and her acting is analogous to the relation between a person’s knowing pain and experiencing pain. Just as knowing pain is unified with experiencing pain, so a person’s (relevant) knowing is unified with her acting.

One might interpret Wang to be saying that a person can know pain some time after she has stopped experiencing it. Such a reading implies that a person who is not currently *φ*-ing anymore can still be said to possess knowledge that is unified with her *φ*-ing. This requires us to construe the unity of A and B in a way that is compatible with saying that you can have A without B, and hence it requires us to explain away those passages that suggest that a person cannot be knowing without acting.

I believe that there is a more straightforward alternative. We can take Wang to be saying that a person is *knowing* pain so long as she is experiencing it, and the moment her pain ceases, she might know what the pain was like, but she is no longer *knowing* pain.[[15]](#footnote-15) Hence, just as knowing pain and experiencing pain occur simultaneously, so (relevantly) knowing and acting occur simultaneously. To Wang’s claim that a person cannot be knowing without acting, we may now add: “without action, there can be no extension of knowledge” (IPL 139; IPL 133). The *Unity Thesis* therefore entails that knowledge-in-action occurs simultaneously with its corresponding action.[[16]](#footnote-16)

There is a simple way to make sense of Wang’s simultaneity claim. A person who is knowing pain knows that the sensation she is currently undergoing is painful. If she didn’t know that her sensation was painful, she wouldn’t be experiencing pain. Analogously, a person who is knowing-to knows to perform what she is currently doing as a way for her to *φ*. If she didn’t know to do *this* to enacta way for her to *φ*, she wouldn’t be intentionally *φ*-ing. Hence, there is a similarity between knowing-to and knowing occurrent sensations – a similarity which, I believe, forms the basis of Wang’s argument.

Wang-exegesis aside, I believe that one reason why the simultaneity claim has wider philosophical relevance is this. It implies that knowing-to is necessary for acting intentionally. If Farmer Zhang does not know to do *this* to politely welcome a governmental envoy, but, unbeknownst to Farmer Zhang, the way he greeted the envoy just turned out to be the polite way of greeting a government official, his lack of knowing-to explains why his otherwise polite greeting was an accident. This also helps explain, for example, why it does not seem possible for you to intentionally win a fair lottery. You cannot know whether drawing this particular number is a way for you to win the lottery, and hence you are not in a position to *know to* draw this particular number to win. Hence, if a person does not know to do *this* for her to *φ*, then even if she does something that accidentally coincides with a way for her to *φ*, she cannot be said to be *φ*-ing intentionally.

**3.3 Objection: Knowing entails being motivated to act**

One might draw on the following passage to raise an objection against the proposal I have defended so far:

The *Great Learning* gives us examples of genuine knowing and acting, saying that it is ‘like loving a beautiful sight or hating a bad odor.’ Seeing a beautiful sight is a case of knowing, while loving a beautiful sight is a case of acting. As soon as one sees that beautiful sight, one naturally loves it. It is not as if you first see it and only then, intentionally, you decide to love it. Smelling a bad odor is a case of knowing, while hating a bad odor is a case of acting. As soon as one smells that bad odor, one naturally hates it. It is not as if you first smell it and only then, intentionally, you decide to hate it. Consider the case of a person with a stuffed-up nose. Even if he sees a malodorous object right in from of him, the smell does not reach him, and so he does not hate it. This is simply not to know the bad odor. (IPL5)

According to IPL 5, Wang Yangming says that experiencing a beautiful sight is relevantly analogous to the “knowing” part of his *Unity Thesis*, and loving a beautiful sight is relevantly analogous to the “acting” part. But loving a beautiful sight is an affective response, one that (presumably) entails having desires. Therefore, the “acting” part of Wang’s thesis entails having desires. If knowing is unified with something that entails having desires, then knowing must entail having desires as well. Neither knowing-that nor knowing-how entails having desires, because you can know that *p* or know how to *φ* without necessarily wanting to perform any of the corresponding actions. According to Huang (2017, 2021, 2022a, 2022b), the only plausible candidate for the kind of knowledge that is necessarily accompanied by a desire is knowledge that you ought to *φ*. Hence, Wang’s point is that there is some knowledge the possession of which necessarily motivates the knower to act. This is, according to some interpreters, what the *Unity Thesis* amounts to: if you are (relevantly) knowing, then you necessarily have the corresponding motivation to act (see, e.g., Ivanhoe 2000: 63-64; 2009: 113; Lee 2023: 116). One might therefore object that my interpretation cannot make sense of the fact that (relevant) knowledge is necessarily action-motivating, and hence that my interpretation is inadequate.

What the objection gets right is that Wang is clearly interested in motivating his students to act virtuously, that being so motivated is characteristic of virtuous people, and that Wang proposes the *Unity Thesis* partly because he believes that accepting it will aid his students in their task of becoming virtuous (e.g., IPL 5, 26). However, the objection fails, because it rests on a misreading.

Consider the following case. Farmer Zhang knows that he ought to save a child when it is about to fall into a well, and suppose that he is also motivated to act accordingly. If you prefer to think that Farmer Zhang has a “moral commitment” (Angle 2009: Ch.7), a “wholehearted inclination” free of motivational conflict (Lederman 2022b: 207-8), or is in some other affective state, feel free to augment the case as you see fit. In situation C, there is a child that is about to fall into a well and Farmer Zhang rushes to save it. In situation not-C, there is no such child, hence Farmer Zhang is not saving any children threatened by wells. In C, Farmer Zhang is (relevantly) knowing and acting. But in not-C, he is (relevantly) knowing and not acting. Nothing about Farmer Zhang’s knowledge of what he ought to do (or his motivation, moral commitment, etc.) has changed between C and not-C, and yet in one case he is acting, but in the other case he is not. Hence, an agent can be (relevantly) knowing without acting. But Wang clearly says that one cannot be knowing without acting. And so the objection fails to make sense of Wang’s *Unity Thesis*.

Someone sympathetic towards the standard proposal might respond in various ways. One way is by saying that Farmer Zhang’s (relevant) knowledge is partly a *disposition* to save others in C, and that a disposition to *φ* in C is unified with *φ*-ing in C. But such a proposal faces the problem that a person can be disposed to *φ* in apt circumstances without ever being in those circumstances. Hence, a person can be disposed to *φ* without ever *φ*-ing. But Wang Yangming is clear that knowing is unified with *acting*. It would be implausibly misleading if Wang had meant to say that a person who is not currently doing anything is nonetheless acting by virtue of being disposed to act.

A second, more promising way to respond is by appealing to mental acts, rather than dispositions. The suggestion is that, although nothing about Farmer Zhang’s knowledge has changed between C and not-C, nothing about his knowledge-relevant acting has changed either. Farmer Zhang is relevantly acting in not-C, when he is not saving children, as much as he is relevantly acting in C, when he is saving children. The reason why Farmer Zhang is relevantly acting in both C and not-C is because in both situations he is *desiring* tosave threatened children, and Wang claims that desiring is itself a mental action (e.g., IPL 132). Hence, on this reading, Wang’s *Unity Thesis* should not be understood as saying that (relevant) knowing is unified with *φ*-ing, but rather that it is unified with the mental act of *desiring* to *φ*.[[17]](#footnote-17) Experiencing a beautiful sight is related to loving a beautiful sight as knowing that I ought to *φ* is related to desiring to *φ*.

One might worry that such a response misleadingly construes the *Unity Thesis* to be about mental acts, rather than overt bodily events. But there is an even deeper problem here. Recall that one feature of Wang’s notion of unity is simultaneity: if X and Y are unified, then I cannot have X without at the same time having Y, and I cannot have Y without at the same time having X. Therefore, if seeing a beautiful sight is unified with loving a beautiful sight, then I cannot have one without at the same time having the other. I take this to mean that a person can love an occurrent sensation only if she is experiencing it. But it is surely false that a person’s knowing that she ought to *φ* is unified in this way with her desire to *φ*. This is because Ican desire to *φ* without simultaneously knowing whether I ought to *φ*. For example, I can be unsure whether I ought to tell my friend a hurtful truth even though I really want to tell her. If this is right, then desiring to *φ* does not entail knowing that I ought to *φ*, and the two do not form a unity*.*[[18]](#footnote-18)

I set the objection aside, but I draw a possible lesson from it. I speculate that, perhaps, commentators on Wang’s *Unity Thesis* have struggled to make sense of the claim that knowing and acting are unified, because they not only share popular assumptions about knowledge but also about action. As I mentioned at the outset of this paper, on a widely endorsed theory of action, what explains that I am intentionally raising my arm is that some purely inner mental states, such as beliefs, desires, or intentions, appropriately set into motion the bodily event of my arm going up. This is, of course, an instance of the *Causal Theory of Action*, which is also known as the “standard story” of intentional action (see Smith 2004; 2010)*.* It implies a view on which intentional actions are composed out of a purely inner mental state and some overt bodily movements. But, of course, the presence of any such mental state does not guarantee that the corresponding bodily event will occur. It is therefore difficult to make sense of the claim that knowing and acting form a unity, given that, at best, knowledge might be able to set a bodily event into motion, but the presence of knowledge on its own, even when augmented by desire, does not guarantee that any action will occur. I speculate that this might be the reason why many commentators on Wang Yangming retreat from saying that knowing and *acting* are unified and instead propose that knowing and *intending* (or *desiring*) to act are unified.

I do not believe that we have to resort to such a roundabout way of making sense of Wang’s *Unity Thesis*. What I see as Wang’s basic idea is that knowledge can be present *in* the action, rather than causing or guiding the action from the outside. This not only has the exegetical advantages I have described above, it is also independently attractive, because it offers us a way of thinking about the nature of action that avoids some long-standing issues in contemporary philosophy of action. In what follows, I focus on one of these issues – namely, the problem of deviant causation.

**4. The *Unity Thesis* and the Problem of Deviant Causation**

An influential way of thinking about intentional action is that, for an action to be intentional, it must appropriately match up with the agent’s intentions, and intentions provide motivation and plans for action (e.g., Bishop 1989; Bratman 1987; Shepherd 2021).

The problem with such a view is that, sometimes, a person can intend to *φ*, succeed at *φ*-ing, and yet not have *φ*-ed intentionally, because the link between the agent’s *φ*-ing and her intention to *φ* is too accidental. In many such cases, the agent’s *φ*-ing does not amount to intentionally *φ*-ing because her action is *deviantly caused* by her intention. A classic example of deviant causation is given by Davidson (1980): a mountaineer intends to murder his companion by letting go of the rope during a trip, but his intention to do so makes him so nervous that, at some point during their climb, he loses his grip and his companion falls to his death. Although his action of letting go of the rope ends up matching his intention, it only does so accidentally. This is because the mountaineer’s intention causes his nervousness, and the nervousness causes him to lose his grip. Therefore, his intention does not cause the companion’s fall in the *right* *way* for the mountaineer’s action to count as intentional.

As it turns out, it is very difficult to give a satisfying account of what that “right way” is supposed to be. Moreover, there is some reason to be skeptical about our prospects. Epistemologists interested in the nature of knowledge have faced a structurally similar problem: if we decompose knowledge into a purely inner mental component (belief) and an external one (the truth of a proposition expressed by the belief), then we must give an account of how these two components relate to one another for them to yield knowledge. The simple fact that epistemologists have failed to offer an explanation that survives (Gettier-style) counterexamples, despite the large amount of highly specialized work that this question has produced over the last sixty years, suggests that the problem might be intractable.[[19]](#footnote-19) But if the problem of deviant causation and the Gettier problem are structurally similar, and the Gettier problem might be intractable, then we have reason to believe that the problem of deviant causation will continue to be a problem in philosophy of action.

One promising way forward, therefore, is to reject the underlying assumption that has led us to causal deviance – and that is the assumption that your bodily movements constitute actions if and only if they are appropriately caused by specific purely inner mental states. If we understand intentional action in terms of knowing-to, we therefore avoid having to specify the right way in which an agent’s purely inner mental states have to match her overt bodily movements, because knowing-to is not a purely inner mental state, nor does it cause a person’s acting. The problem of deviant causation does not arise.

In response to my proposal, one might object that non-causal accounts (such as the one I am proposing) face their own worries about causal deviance. Sarah Paul (2011) draws attention to one such deviance problem, which she develops specifically for Anscombean non-causal approaches – namely, the problem of *deviant formal causation*.

According to Anscombe’s (1957) influential proposal, a person is acting intentionally just in case she knows what she is doing as she is doing it. But Paul argues that such a view has difficulty distinguishing an intentional action from the foreseen *side-effects* of the action. Adapting one of Anscombe’s own examples, Paul suggests that we can imagine a case of two gardeners, both of which know that they are currently pumping poisoned water into a cistern and that doing so will (i) get them paid for the job and (ii) poison the inhabitants. Gardener A finds poisoning the inhabitants undesirable and pumps poisoned water for the sake of getting paid. For him, poisoning the inhabitants is not his aim but a foreseen side-effect of his act of pumping water into the cistern. But Gardener B aims to poison the inhabitants and pumps water for the sake of doing so. For him, getting paid is but a welcome side-effect. Intuitively, the actions performed by Gardener A and Gardener B are not identical. And yet, the two gardeners are in the same epistemic state: they both know, in the same way, what they are doing as they are doing it. This is a problem for an Anscombean non-causal account of intentional action.

But it is not a problem for Wang Yangming. Reading Wang’s *Unity Thesis* in terms of knowing-to allows us to avoid the problem of deviant formal causation, because we can reject the claim that Gardener A and B are in the same epistemic state, while granting Paul’s (2011) argument that both of them have the same propositional knowledge of what they are doing. The reason why they epistemically differ is because they differ in theirknowing-to.Gardener A knows to do *this* to get paid, while Gardener B knows to do *this* to poison the inhabitants. Presumably, if Gardener A were to be unsure whether doing *this* will get him paid, he would change course, whereas Gardener B would not.[[20]](#footnote-20)

Insofar as we can explain acting intentionally in terms of the enactment of knowledge, and we can explain the enactment of knowledge in terms of knowing-to, I believe that we are offered an analysis of intentional action that avoids the problem of deviant causation. Although I cannot offer a complete defense of such a view here, I hope to have shown that Wang’s *Unity Thesis* can not only be given an exegetically fitting interpretation, it also has the potential to offer us a plausible alternative way of thinking about the nature of intentional action.[[21]](#footnote-21)

**Bibliography**

Angle, Stephen C. 2009. *Sagehood: The Contemporary Significance of Neo-Confucian Philosophy*. New York: Oxford University Press.

Angle, Stephen C., and Justin Tiwald. 2017. *Neo-Confucianism: A Philosophical Introduction*. Malden, MA: Polity.

Anscombe, G.E.M. 1957. *Intention*. Cambridge, MA: Harvard University Press.

Bishop, John C. 1989. *Natural Agency: An Essay on the Causal Theory of Action*. Cambridge: Cambridge University Press.

Bratman, Michael. 1987. *Intention, plans and practical reason*. Cambridge, MA: Harvard University Press.

Brys, Waldemar. 2022. “The Epistemology of Mengzian Extension.” In *Knowers and Knowledge in East-West Philosophy*, edited by Karyn L. Lai, 43-61. Cham: Palgrave Macmillan.

Brys, Waldemar. 2023. “Epistemology in the *Mencius*.” In *Dao Companion to the Philosophy of Mencius*, edited by Yang Xiao and Kim-chong Chong, 491-514. Cham: Springer.

Chan, Wing-tsit 陳榮捷. 1963. *Instructions for Practical Living and Other Neo-Confucian Writings by Wang Yangming*. New York: Columbia University Press.

Chang, Tzu-li 张子立. 2016. “Re-exploring Wang Yangming’s Theory of “Liangzhi”: Translation, Transliteration, and Interpretation.” *Philosophy East and West*, 66: 1196-1217.

Chen, Lai 陳來. 1991. *The Realm of Being and Non-being: The Spirit of Wang Yangming’s Philosophy* 有無之境: 王陽明哲學的精神. Beijing: Renmin Chubanshe人民出版社.

Cua, Antonio S. 1982. *The Unity of Knowledge and Action: A Study in Wang Yang-ming’s Moral Psychology*. Honolulu: University Press of Hawai’i.

Davidson, Donald. 1980. “Actions, Reasons and Causes.” In *Essays on Actions and Events*, 3–21. New York: Oxford University Press.

Farkas, Katalin. 2017. “Practical Know-Wh.” *Nous* 51: 855-870.

Frisina, Warren G. 2002. *The Unity of Knowledge and Action*. Albany: State University of New York Press

Hetherington, Stephen. 2011a. *How to Know: A Practicalist Conception of Knowledge*. Malden, MA: Wiley-Blackwell.

Hetherington, Stephen. 2011b. “Knowledge and Knowing: Ability and Manifestation.” In *Conceptions of Knowledge*, edited by Stefan Tolksdorf, 73-100. Berlin: De Gruyter.

Hetherington, Stephen. 2016. *Knowledge and the Gettier Problem*. Cambridge: Cambridge University Press.

Hetherington, Stephen. 2017. “Knowledge as Potential for Action.” *European Journal of Pragmatism and American Philosophy* 9: 1-19.

Hetherington, Stephen. 2021a. “Knowledge as skill.” In *The Routledge Handbook of Philosophy of Skill and Expertise*, edited by Ellen Fridland and Carlotta Pavese, 168-178. New York: Routledge.

Hetherington, Stephen. 2022a. “Knowing-to.” In *Knowers and Knowledge in East-West Philosophy*, edited by Karyn L. Lai, 17-41. Cham: Palgrave Macmillan.

Hetherington, Stephen. 2022b. *Defining Knowledge*: *Method and Metaphysics*. Cambridge: Cambridge University Press.

Huang, Yong. 2017. “Knowing-that, Knowing-how, or Knowing-to?” *Journal of Philosophical Research* 42: 65–94.

Huang, Yong. 2021. “Belief, Desire, and Besire: Slote and Wang Yangming on Moral Motivation.” In *Michael Slote Encountering Chinese Philosophy*, edited by Huang Yong, 81-98. London: Bloomsbury Academic.

Huang, Yong. 2022a. “A Third Platonic Problem for Sosa? Or How Wang Yangming Can Know Better than *Full Well*?” In *Ernest Sosa Encountering Chinese Philosophy*, edited by Huang Yong, 203-220. London: Bloomsbury Academic.

Huang, Yong. 2022b. “Tu Wei-ming’s *Tizhi* and the Confucian Contribution to Contemporary Epistemology.” *Philosophy East and West* 72: 739-757.

Hyman, John. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.

Ivanhoe, Philip J. 2000. *Confucian Moral Self Cultivation*, 2nd. ed. Indianapolis: Hackett.

Ivanhoe, Philip J. 2009. *Readings from the Lu-Wang School of Neo-Confucianism*. Indianapolis: Hackett.

Ivanhoe, Philip J. 2022. “The Introspective, Perceptual, and Spontaneous Response Models of Wang Yangming's Philosophy.” *Journal of Confucian Philosophy and Culture* 38: 43-65.

Lederman, Harvey. 2022a. “What Is the “Unity” in the “Unity of Knowledge and Action”?” *Dao* 21: 569-603.

Lederman, Harvey. 2022b. “The Introspective Model of Genuine Knowledge in Wang Yangming.” *Philosophical Review* 131: 169-213.

Lederman, Harvey. 2023. “Conceptions of Genuine Knowledge in Wang Yangming.” In *Oxford Studies in Epistemology* *7*, edited by Tamar Szabó Gendler, John Hawthorne, and Julianne Chung, 134-175. Oxford: Oxford University Press.

Lee, Ming-huei 李明輝. 2023. “Wang Yangming’s Doctrine of the “Unity of Knowing and Acting” in the Light of Kant’s Practical Philosophy.” *Journal of Confucian Philosophy and Culture* 39: 91-128.

Lu, Yinghua 卢盈华. 2019. “Wang Yangming’s Theory of the Unity of Knowledge and Action Revisited.” *Philosophy East and West* 69: 197-214.

Paul, Sarah K. 2011. “Deviant Formal Causation.” *Journal of Ethics & Social Philosophy* 5.3: 1-24

Sayre, Kenneth M. 1997. *Belief and Knowledge: Mapping the Cognitive Landscape*. Lanham, MD: Rowman & Littlefield.

Shepherd, Joshua. 2013. “The Contours of Control.” *Philosophical Studies* 170: 395–411.

Shepherd, Joshua. 2021. *The Shape of Agency: Control, Action, Skill, Knowledge*. Oxford: Oxford University Press.

Slingerland, Edward. 2019. *Mind and Body in Early China*. New York: Oxford University Press.

Smith, Michael. 2004. “The Structure of Orthonomy.” *Royal Institute of Philosophy Supplement* 55: 165-193.

Smith, Michael. 2010. “The Standard Story of Action.” In *Causing Human Actions*, edited by Jesús Aguilar and Andrei A. Buckareff, 45–56. Cambridge, MA: The MIT Press.

Tiwald, Justin, and Bryan W. Van Norden. 2014. *Readings in Later Chinese Philosophy*. Indianapolis: Hackett.

Valaris, Markos. 2020a. “Knowing what you are doing: Action-demonstratives in unreflective action.” *Ratio* 33: 97–105.

Valaris, Markos. 2020b. “Thinking by doing: Rylean regress and the metaphysics of action.” *Synthese* 197: 3395-3412.

Valaris, Markos. 2022a. “Knowledge Out of Control.” *The Philosophical Quarterly* 72: 733-753.

Valaris, Markos. 2022b. “Control and Knowledge in Action: Developing Some Themes from McDowell.” In *Reason in Nature: New Essays on Themes from John McDowell*, edited by Matthew Boyle and Evgenia Mylonaki, 153-170. Cambridge, MA: Harvard University Press.

Valaris, Markos. Manuscript. *Knowledge in Action: A Knowledge-First Theory of Action*.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. New York: Oxford University Press.

Yang, Xiaomei. 2009. “How to Make Sense of the Claim 'True Knowledge is What Constitutes Action'.” *Dao* 8: 173-188.

Zheng, Zemian 鄭澤綿. 2018. “An Alternative Way of Confucian Sincerity: Wang Yangming's "Unity of Knowing and Doing" as a Response to Zhu Xi's Puzzle of Self-Deception.” *Philosophy East and West* 68: 1345-1368.

Zheng, Zemian 鄭澤綿. 2023. “Mencius and Wang Yangming.” In *Dao Companion to the Philosophy of Mencius*, edited by Yang Xiao, Kim-chong Chong, 201-218. Cham: Springer.

Zhang, Xuezhi 張學智. 2021. *History of Chinese Philosophy in the Ming Dynasty*. Singapore: Springer.

1. I adopt this usage from Valaris (2022a). Henceforth, I bracket the question of whether mental states are “purely inner” in some other way, e.g., in the sense of being vaguely located “in the mind” (see Slingerland 2019 for a discussion on whether early Chinese philosophers employed such a notion of psychological interiority). [↑](#footnote-ref-1)
2. Hence, I take Wang, in effect, to agree with Anscombe (1957) on the special role that knowledge plays in explaining intentional action. However, I do not take Wang to be committed to a broadly Anscombian account of practical knowledge. I return to this in Section 4. [↑](#footnote-ref-2)
3. This does not mean (nor do I intend to argue) that *only* intentional actions are unified with knowledge. It is likely that, for Wang, some non-intentional actions (specifically mental actions) are also unified with knowledge. I bracket such a possibility here. My claim is that, according to the *Unity Thesis*, knowledge is unified with all but perhaps not only intentional actions. [↑](#footnote-ref-3)
4. For a good analysis of how these terms are related, see Lederman (2022a). Unless otherwise specified, all translations of the *Chuánxí Lù* 傳習錄(*Instructions for Practical Living*; hereafterIPL), which isour primary source for Wang’s views, are adapted from Wing-tsit Chan (1963) and, where possible, from Ivanhoe (2009) and Ivanhoe’s revised translations published in Tiwald & Van Norden (2014). [↑](#footnote-ref-4)
5. For discussion, see Angle & Tiwald 2017: Ch. 6. [↑](#footnote-ref-5)
6. See, for example, Cua (1982), Yang (2009), Lee (2023: 111). For the view that the *Unity Thesis* applies not only to morally virtuous actions, see, e.g., Frisina (2002: Ch.4), Zhang (2021: 135). [↑](#footnote-ref-6)
7. Some have suggested that this passage is evidence that only knowledge-how is unified with acting, because the quality of your knowing how to *φ* is improved with the quality of your reliable *φ*-ing (Lu 2019: 206-207). The problem with such a view is that a person can know how to *φ* without currently *φ*-ing, and hence that the two are not relevantly unified. I return to this in Section 3. [↑](#footnote-ref-7)
8. Here, I follow Ivanhoe (2009) and Angle & Tiwald (2017) in translating *zhǔ yì* 主意 as “intent”. However, I take my point to be compatible with Lederman’s (2022a: 593n28) suggestion that *zhǔ yì* 主意 is to be translated as “main goal”. This is because, if your *φ*-ing is done for the sake of achieving some goal, then you are *φ*-ing for a reason, and *φ*-ing for reasons has long been treated as a characteristic feature of intentional action (see, e.g., Davidson 1980). [↑](#footnote-ref-8)
9. For discussion and defense of the concept of knowing-to as I am using it here, see Hetherington (2022a). For an argument that knowing-to also features in the 3rd-century BCE text *Mengzi*, see Brys (2022; 2023), and for the argument that Wang Yangming was heavily influenced by the *Mengzi*, see, e.g., Angle & Tiwald (2017); Zheng (2023). [↑](#footnote-ref-9)
10. Huang Yong also calls the relevant kind of knowledge “knowing-to”. For ease of reference, I refer to Huang’s notion as “knowing that one ought to *φ*” rather than“knowing-to.” [↑](#footnote-ref-10)
11. This has puzzled commentators, with some arguing that, if knowledge-in-action is the beginning of an action, then the relevant knowledge must temporally precede the corresponding action (e.g., Chen 1991: 99; Lee 2023: 103). But the metaphysics here is incorrect. The beginning of *φ*-ing is not an event separate from *φ*-ing. Hence, if knowledge-in-action is the beginning of *φ*-ing, thenit is part of *φ*-ing – it does not temporally precede it. For textual reasons to resist the temporal reading, see also Lederman (2022a: 594-5) and Huang (2017: 73-5). [↑](#footnote-ref-11)
12. Farkas (2017) makes a similar point in her argument that some answers to wh-questions are only available during the performance of an action, but she takes knowledge of such answers to be propositional. Nothing I say here commits me to the view that knowing-to is *propositional* knowledge. [↑](#footnote-ref-12)
13. Additionally, this is one reason why knowing-to differs from knowing-that and knowing-how. I can know that saying *these* words is hurtful, and hence not do it. And I can know how to say hurtful words without putting my know-how into practice. But I cannot know to dowhat I am currently doing withoutcurrently doing it. This is because knowledge is factive: if you know that *p*, it cannot be that *p* is false. Hence, if you are knowing to do *this* for you to *φ*, then it cannot be that doing *this* is not a way for you to *φ*. I thank an anonymous referee for pressing this point. [↑](#footnote-ref-13)
14. Hetherington (2011b; 2017) argues for conceptually distinguishing a person’s having knowledge that *p* from her *knowing* that *p*. He conceives of the former as a state that admits of manifestation and the latter as a manifesting of the relevant state in action. On such a view, all cases of knowing that *p* are *p*-related actions. I take this to be compatible with the proposed claim that, to get from a capacity for *φ*-ing to the corresponding act of *φ*-ing, the agent must know to do *this* for her to *φ*. [↑](#footnote-ref-14)
15. Chang (2016: 1208) and Lederman (2022a: 589-90) offer a reading along these lines. [↑](#footnote-ref-15)
16. Although my argument emphasizes that, for Wang, (relevant) knowing and acting occur at the same time, this is compatible with Lederman's(2022a) explication of the *Unity Thesis* in biconditional form: a person is (relevantly) knowing *if and only if* she is (correspondingly) acting. [↑](#footnote-ref-16)
17. Taking the “action” part of Wang’s *Unity Thesis* to only refer to mental acts, such as desiring, appreciating, intending, etc., is a frequent interpretative move (see, e.g., Zheng 2018: 1355; Huang 2017: 78; but for an important exception, see Yang 2009). Huang Yong sometimes explicitly glosses the *Unity Thesis* in terms of the unity of “knowledge and (desire for) action” (e.g., Huang 2021: 90). [↑](#footnote-ref-17)
18. Might it help if we replace “desiring to *φ*” by “appropriately responding to the details of the situation”, as recently proposed by Ivanhoe (2022: 61)? I do not think so. If telling my friend a hurtful truth is the appropriate response in a specific situation, then surely I can appropriately respond by accident – namely, I can blurt out the hurtful truth without (relevantly) knowing that this is the appropriate thing to do. What if we draw on recent work by Lederman (2022b; 2023) and argue that not *all* desires are unified with knowing, but only *wholehearted* desires? Although this is not how Lederman would himself conceive of the *Unity Thesis* (and a discussion of his views deserves more space than I can offer here), one might nonetheless be tempted to draw inspiration from his account of wholehearted desires in defense of the view that desiring to *φ* entails (relevant) knowing.Very roughly, according to Lederman (2022b: 191n36), a desire is wholehearted *iff* there is no other desire that conflicts with it. All cases of (relevantly) knowing are cases of my faculty of *liángzhī* 良知 knowing that an occurrent desire is good (*ibid*.: 197). Hence, one might suggest that I cannot wholeheartedly desire to *φ* without (relevantly) knowing. The problem with such a suggestion, it seems to me, is that there can be many actions and corresponding desires on which a person’s *liángzhī* is silent. Lederman (2022b: 178-90) admits that *liángzhī* knows that “particular mental events are right or wrong, *if they are*.” (My own emphasis) But if some mental events are neither right nor wrong, it is possible that I can have a wholehearted desire for something on which my *liángzhī* is silent, and hence I can have a wholehearted desire without (relevantly) knowing. In response, one might restrict the scope of the *Unity Thesis* to only concern morally virtuous mental actions. That is, if we say that the claim “all wholehearted desires are unified with knowing” only applies to cases where wholehearted desiring is a morally virtuous mental action, then perhaps we might avoid the issue that, for Wang, a person can have a wholehearted desire without (relevantly) knowing, because all morally virtuous actions necessarily involve (relevant) knowing. The problem with such a response is that it would require us to restrict the *Unity Thesis* to specifically *mental* actions, and there is no evidence, as far as I can tell, that Wang holds such a restrictive view. [↑](#footnote-ref-18)
19. For a detailed analysis of our failure to solve the Gettier problem, see Hetherington (2016). In recent years, a growing trend has been to develop approaches that aim to *avoid* the Gettier problem, rather than to *solve* it: for example, analyses of knowledge in terms of practical ability (Hetherington 2011a; 2021b; Hyman 2015), true belief (Hetherington 2022a), or views that take knowledge to be unanalysable (Sayre 1997: Ch.5; Williamson 2000). [↑](#footnote-ref-19)
20. For a neo-Anscombean answer to the problem, one that rests on rejecting Sarah Paul's claim that the two agents have the same kind of knowledge of what they are doing, see Valaris (ms: Ch.5). [↑](#footnote-ref-20)
21. I have greatly benefitted from written comments on previous drafts of this paper by Stephen Hetherington, Justin Tiwald, an anonymous referee, and especially Karyn Lai and Philippe Brunozzi. I also thank Markos Valaris and Yuchen Guo for discussion. [↑](#footnote-ref-21)