The Introspective Model of the Unity of Knowledge and Action

Harvey Lederman

May 30, 2019
Please do not cite without permission, but comments welcome!

Abstract

This paper presents a new interpretation of the great Ming dynasty philosopher Wang Yangming's (1472-1529) doctrine of the "unity of knowledge and action" (知行合一). My interpretation focuses on a new understanding of Wang's notion of "genuine knowledge", according to which Wang held that genuine knowledge requires freedom from a particular form of doxastic conflict. I argue that Wang believed that freedom from this form of doxastic conflict requires freedom from motivational conflict, and that if a person is free from motivational conflict, they are acting virtuously. Thus, on my interpretation, Wang held that knowledge and action are unified at least in the sense that, if a person has genuine knowledge, they are acting virtuously.

1 Introduction

In 1506, in exile in Longchang in Guiyang, Wang Shouren (王守仁, Yangming 陽明 1472-1529) experienced a "great enlightenment" (大悟), when a voice seemed to call out to him in the night.[1] A year later, Wang distilled this dramatic revelation in the doctrine of the "unity of knowledge and action" (*zhi xing he yi* 知行合一).[2] The doctrine would come to be seen as one of the major achievements of Ming (明) dynasty (1368-1644) thought, and indeed, of the whole tradition now called "Confucian". It was a central component of the distinctive philosophical outlook which would earn Wang a place on the standard list of the four most important thinkers in this tradition ("孔孟朱王"),

[1] Here I follow the account of Qian Dehong in Wang's *nian pu*; text from Wu *et al.* (2011, 1354-5).

[2] Throughout this paper, I will follow tradition in translating 合一 as "unity" in this slogan. The expression can also mean something weaker, more like "correspondence". I discuss this issue further at the end of Section 2.

alongside Confucius, Mencius, and the twelfth-century scholastic system-builder Zhu Xi (朱熹, Yuanhui 元晦, 1130-1200).

But while the unity of knowledge and action has been widely celebrated, it has also been widely condemned. Wang's contemporaries presented a series of criticisms and putative counterexamples to what they understood by the "unity of knowledge and action". In his replies to these challenges, Wang retreats to the claim that action is unified only with what he calls "genuine knowledge". But rebranding does not a discovery make. Someone who claims to have discovered an elixir which guarantees eternal life, but in the small print states that by "eternal life" they mean "ordinary life", does not by their verbal alchemy transform the placebo they possess. Similarly, if "genuine knowledge" is defined as "that which is unified with action", then Wang's doctrine might be true, but it would be trivially so. So the "unity of knowledge and action" has come to seem a misleading advertisement for a triviality.

In this paper I present a new interpretation of the unity of knowledge and action, which shows that this criticism is mistaken. I argue that Wang characterizes genuine knowledge as an elevated form *of knowledge*, independently of its relationship to action. On my interpretation, Wang holds that a person has genuine knowledge if and only if they are free from a particular form of doxastic conflict. Wang then advances substantive theses connecting freedom from this form of doxastic conflict to virtuous action. The result is that Wang did not seek to pawn off placebo under false pretences. Instead, under the heading of the unity of knowledge and action, he advanced a series of striking claims relating freedom from doxastic conflict to freedom from motivational conflict, and relating freedom from motivational conflict, in turn, to virtuous action.

I will argue throughout this paper that Wang held systematic, detailed views in moral psychology and moral epistemology. This aspect of my interpretation is significant in its own right. Wang has often been described as a visionary thinker, but dismissed as one who was not capable of or not interested in advancing systematic theories. But I believe this assumption has simply been a self-supporting prejudice. Since it was thought to be hopeless to attempt to find a systematic theory in Wang's writings, scholars have not attempted to find one there; a bias towards high-level discussions of Wang's views in the scholarship has then seemed to confirm the prejudice that his views resist systematic exposition. This paper challenges that prejudice. I hope that – whatever readers may make of the details of the interpretation I present – the paper will persuade others that Wang's writings richly repay careful, systematic examination.

The development of my interpretation and the detailed discussion of how it is supported by the texts has required somewhat lengthy exposition. As a result, I have been

unable to include thematic discussion of others' interpretations of the unity of knowledge and action in this paper, although I document differences in their interpretations of particular passages throughout. A companion paper, "Perception and Genuine Knowledge in Wang Yangming", develops alternative interpretations and discusses my disagreements with other scholars in detail. Both papers are intended to be self-contained, but specialists may wish to read them together.[3]

2 The unity of knowledge and action

This introductory section offers a close reading of Wang's most famous discussion of the unity of knowledge and action. I argue that Wang under this heading espoused a precise claim about the connection between what he called "genuine knowledge" and virtuous action. In particular, I will argue that for a range of virtues he endorsed claims of the form:

Filiality Unity A person genuinely knows filial piety if and only if they are acting filially.

The section sets the stage for the main body of the paper, which is dedicated to providing a detailed analysis of the psychological mechanisms which Wang believed underwrote the truth of Filiality Unity (and related claims for other virtues). The end of the section provides a more detailed outline of the remainder of the paper, once the key terms in this thesis have been introduced.

Before I turn to the text which will occupy us throughout this section, I want to introduce the the word *zhi*, which I've translated as "knowledge" in the slogan "the unity of knowledge and action". In addition to being the noun for "knowledge", the word can also be used as the verb "know". But the word also has a use in which it diverges from the cognates of the English word "know", and in which it describes a conscious episode. To a first approximation we might render the word in this use by "is conscious of" or even "apprehend". In one passage, Wang argues against his interlocutor's proposal

[3]This paper will be limited in a further way, as well: I will focus almost exclusively on Wang's views in moral psychology and moral epistemology. As many authors have observed, Wang's views in these areas are motivated by his broader metaphysics and anthropology, for instance, views that he expresses by saying that the sage and the world around them share one body. Some might worry that setting aside Wang's metaphysics leads inevitably to a distortion of his views, but I hope that we can make some progress on his moral philosophy without a full interpretation of his metaphysics before us, and that what I say below is compatible with many reasonable ways of understanding his broader metaphysics. But I certainly agree that a fuller account of Wang's thought than I can offer here (and one I hope to offer in the future) would address these connections. For discussion, see e.g. Frisina (1989), Cua (1993), Tien (2010), Tien (2012), and especially now Ivanhoe (2018).

that people do not *zhi* when they are asleep, by observing that people can be awakened by sensory stimulation.[4] Since of course people can know things while asleep – one doesn't have to check whether someone is awake to determine whether they know that Luoyang was the capital of Tang – it is natural to see Wang and his interlocutor as debating whether people are conscious of or apprehend features of their environment while they are asleep, rather than whether they know anything while asleep. There aren't perfect English correlates for this use of the word; whereas the relevant kind of apprehension is instantaneous (it is similar in this respect to recognition), an episode of *zhi*-ing can last for longer than an instant; whereas being conscious of something does not require a particularly high degree of cognitive accomplishment *zhi*-ing typically involves a significant cognitive accomplishment (although this feature of the expression is not on display in the passage just cited about sleep). The situation is further complicated by the fact that *zhi* in this alternative use retains its connection to its other meaning, in which it is close in meaning to the word "know". In light of these complexities, throughout the paper I will continue to translate the word *zhi* as "know", but in my interpretations I'll often speak of "episodes of knowledge" and "apprehending" (which I use to mean "experiencing an episode of knowledge"), when it is clear Wang has the alternative meaning in mind. These locutions are intended merely as suggestive placeholders which indicate that I suspect Wang is using the word to describe an event whose temporal profile differs from knowledge but which is nevertheless for him closely related in significance to knowledge. I don't expect the reader to be able to work out what an episode of knowledge is merely from these expressions alone; Wang's idea should become clearer as we encounter further passages in which he discusses it.[5]

We now turn to the text. In the first volume of the main selection of Wang's sayings and writings, the *Instructions for Practical Living* (hereafter, *IPL*, 傳習錄), Xu Ai (徐愛, Riren 曰仁 1487-1517), the transcriber of this section of the work, asks Wang about an apparent difficulty with the doctrine of the unity of knowledge and action:

[4] *IPL* 267 (*QJ* 120). Passages from the *Instructions for Practical Living* are first cited by the section number of Chan's editions (Chan (1963), Chan (陳榮捷) (1983), abbreviated "*IPL*"), for ease of reference for those without Chinese. All quotations are followed by a page number either in Wu *et al.* (2011) (indicated by "*QJ*") or in Shu & Zha (2016) (indicated by "*QJBB*").

[5] It is received wisdom that almost all classical Indian epistemology focused on a mental *event* (see Matilal (1986, Ch. 1.4 and Ch. 4) and now Perrett (2016, Ch. 2)). The idea is also not alien to the history of Western epistemology. The Stoics focused on events related to belief and knowledge instead of belief and knowledge themselves (Brennan, 2005, p. 65, 69-70). The claim often translated as "thought thinks itself" (αὐτὸν δὲ νοεῖ ὁ νοῦς 1072b19-20) in the discussion of god in Aristotle's *Metaphysics* Λ.7, is naturally understood to describe the mind actively considering itself in a way that requires a fairly demanding form of cognitive grasp. Picking up on Aristotle's Greek usage, some scholastic discussions of *intelligere* have a similar flavor, e.g. Aquinas's discussion of whether angels (unlike people) can *intelligere* many things at the same time (*Summa Theologiae*, I.58.2).

如今人儘有知得父當孝、兄當弟者，却不能孝、不能弟，便是知與行分明是兩件。

For instance, today everyone knows that they should be filial to their parents, and that they should be respectful to their older brothers, but they are unable to be filial, and unable to be respectful. So in this case, knowledge and action are separated, and are clearly two things. (*IPL* 5, *QJ* 4)[6]

Xu's example is related to examples of *akrasia*; the people he describes know that they ought to perform an action, but they voluntarily fail to do it nevertheless. Xu takes this kind of case to threaten the unity of knowledge and action, presumably because the actions of the people are not responsive to what they know about virtuous action. In his reply, Wang defends the doctrine by introducing a distinction between kinds of knowledge:

[T1] 此已被私慾隔斷，不是知行的本體了。未有知而不行者。知而不行，只是未知。聖賢教人知行，正是安復那本體，不是着你只恁的便罷。故《大學》指個真知行與人看，說『如好好色，如惡惡臭』。見好色屬知，好好色屬行。只見那好色時已自好了，不是見了後又立個心去好。聞惡臭屬知，惡惡臭屬行。只聞那惡臭時已自惡了，不是聞了後別立個心去惡。如鼻塞人雖見惡臭在前，鼻中不曾聞得，便亦不甚惡，亦只是不曾知臭。

In this case, knowledge and action have already been divided by selfish desires; they are no longer the original substance (*ben ti*) of knowledge and action.[7] No one has ever known but failed to act. If one knows but does not act, one simply does not yet know. The sages and worthies' teaching for people about knowledge and action, was to stabilize and restore their original substance, not just to do any old thing.

The *Great Learning* points to genuine knowledge and action for people to see.

[6]Translations are mine, although I have always consulted Chan (1963) for selections from the *Instructions for Practical Living* and Ivanhoe (2009) for passages translated there. In the main text I will not mention issues about the dating of various passages, and how the passages might fit into the development of Wang's view. Where a known change of view or emphasis on Wang's part matters to the central argument I'll discuss this in the footnotes.

[7]I will use the traditional translation "original substance" for *ben ti* 本體 throughout the paper. The term has troubled translators for some time, though its meaning in contexts like this one is fairly clear. It describes the condition something would be in if no extrinsic factors interfered with it. The idea was that the state something would manifest without external interference was also indicative of its nature in some deeper sense, which determined its overall behavior in other conditions as well. In this regard the expression describes something similar to what some authors call "essence", but unlike essential properties, a thing can (and things typically do) fail to have properties its "original substance" has. Mou Zongsan (1972) argues that the expression means "in-itself" (so that the present phrase would be "knowledge and action in themselves"; see also the English translation of part of the article as Mou (1973)). Recently scholars have used "inherent reality" (Angle & Tiwald (2017)). Perhaps "original intrinsic condition" would be better than these alternatives; *ti* 體 – which has a very similar meaning on its own – could then be rendered "intrinsic condition". Since all of these translations are jargon which require further explanation to be comprehensible to the reader, it seemed better on balance to stick with tradition.

> It says they are "like loving lovely colors and hating hateful odors." Seeing a lovely color belongs to knowledge, while loving a lovely color belongs to action. But once someone sees a lovely color, he already loves it. It is not that after seeing it he additionally makes up his mind to love it. Smelling a hateful odor belongs to knowledge, while hating a hateful odor belongs to action. But once someone smells a hateful odor, he already hates it. It is not that after smelling it he separately makes up his mind to hate it. It's like a person with his nose blocked: even if he sees something with a hateful smell in front of him, in his nose, he has not smelt it. So while he doesn't really hate it, this is only because he does not yet know the odor. (*IPL*, 5, *QJ*, 4)

Wang says that the knowledge and action in Xu's example are no longer knowledge and action "in their original substance". He goes on to describe this form of knowledge as "genuine knowledge". The expression "genuine knowledge" occurs as early as the *Zhuangzi* (3rd c. BCE) (*The Great and Most Honored Master*, 1), but Wang's usage derives more proximately from Song (宋 960-1279) authors, who distinguished ordinary knowledge from genuine knowledge. In a famous passage, Cheng Yi (程頤, Yichuan 伊川, 1033-1107), for example, says that someone who has previously been mauled by a tiger will blanche with fear at the news that a tiger is roaming the countryside, whereas people who have never encountered a tiger may know that tigers are to be feared, but they will not experience fear. Cheng describes this difference in people's response as illustrative of the difference between genuine knowledge (真知) and ordinary knowledge (常知) (*Ercheng Yishu*, 3.23).[8]

Wang never distinguishes genuine knowledge from "ordinary knowledge" as clearly as Cheng Yi does in these passages. But it is fairly clear that he follows Cheng Yi (and Zhu Xi (朱熹, Yuanhui 元晦, 1130-1200)) in using the expression "genuine knowledge" as a technical term.[9] In our passage, Wang seems to take "knowledge in its original substance" to be equivalent to "genuine knowledge". And although in his first remarks Wang does use "know" without any qualification, it is clear that he has genuine knowl-

[8] For further English language discussion of these precedents see Shun (2010, p. 188), Angle (2018, p. 166). Cf. Yong (2015)

[9] See *IPL* 125, *QJ* 42; *IPL* 133, *QJ* 47-8. Moreover, it is clear in many other passages that Wang thinks there is a distinction in kinds of knowledge, between knowledge that is not extended (致) and knowledge that is. For instance, in *IPL* 138, *QJ* 55 (which follows immediately the passage quoted below as **[T7]**) we find: "Knowing how to perform the details of warming and cooling [one's parents], or knowing how to perform the rituals of serving and nourishing [them] is what is called knowing. But it can not yet be called extended knowing. One must extend this knowing how to perform the details of warming and cooling [one's parents], and actually use it to warm or cool. One must extend the knowledge of how to perform the rituals of serving and nourishing [them], and actually use it to serve and nourish. Only then can it be called extended knowledge." (知如何而為溫凊之節，知如何而為奉養之宜者，所謂知也，而未可謂之致知。必致其知如何為溫凊之節者之知，而實以之溫凊，致其知如何為奉養之宜者之知，而實以之奉養，然後謂之致知。) For similar remarks see the letter to Zhu Yangbo, *QJ* 309.

edge (and not just ordinary knowledge) in mind throughout. He does not intend to claim that the people in Xu's example have no knowledge *at all*; he is only denying that they have genuine knowledge in this technical sense. In fact, he seems to accept Xu's claim that they have knowledge in some lesser sense. The first move in his response to Xu Ai is thus to clarify his thesis further. The thesis is not to be understood as saying that all knowledge and action are unified, but only that knowledge and action in their original substance are. And this knowledge in its original substance is genuine knowledge.

Wang cites two examples from the *Great Learning* to illustrate the relationship between genuine knowledge and action. Loving a lovely color begins no later than seeing it; hating a hateful odor begins no later than smelling it.[10] At the end of the passage Wang discusses someone who can see the source of an odor but cannot smell the odor because his nose is blocked. This discussion suggests that smelling the odor is the only way of knowing it in the sense in which Wang has in mind. One would expect that Wang would similarly hold that seeing a color is the only way of knowing it in the relevant way, so that if one knows a color, one has seen it, and if one knows an odor one has smelled it. Wang furthermore says that just as seeing and smelling "belong" to knowledge, hating and loving "belong" to action. It is fairly clear that he means that seeing and smelling are ways of knowing or forms of knowledge, and that loving and hating are similarly ways of acting or forms of action.[11] Thus, putting all of these claims together: if one knows a beautiful color, one has seen it, and hence loved it, i.e. acted towards it. Similarly, if one knows a bad odor, one has smelled it, and hence hated it, i.e. acted towards it.

Wang's claim that loving and hating are to be considered actions is surprising. One might naturally describe loving and hating as feelings or emotions, but not as actions. It might be that Wang is merely making this claim in the service of mapping the example from the *Great Learning* on to the psychological phenomenon that most interests him.

[10] I have translated the quotation from the *Great Learning* as "loving a lovely color and hating a hateful odor" in an attempt to simulate the fact that the verb "love" is written with the same character (好) as the adjective "lovely" (although they are pronounced differently). Similarly, the verb "hate" is written with the same character (惡) as the adjective "hateful" (although again they are pronounced differently).

[11] If one took "belong to" *shu yu* 屬於 to mean "is a part of", Wang would say here that seeing is a part of knowing, and loving is a part of acting. The latter of these claims especially might seem to fit nicely with what Wang says in *IPL* 132 (*QJ* 46-7), where he describes certain desires as "the beginning of action" 行之始. But the interpretation should be rejected in our passage, because it makes a mess of Wang's line of thought here. Wang argues from the claim that if one sees, one loves to the claim that knowing and acting are tightly connected. If seeing and loving were proper parts of knowing and acting, and didn't suffice for knowing and acting on their own, then the connection between seeing and loving wouldn't suffice to establish an interesting connection between knowing and acting.

Or it might be that Wang himself thought that psychological or affective responses of this kinds are actions in their own right.[12] I will remain officially neutral in this paper on how we settle this question; I won't be assuming that Wang did or did not take loving and hating to be actions.

In the text immediately following **[T1]**, Wang gives examples of the connection between knowledge and action. His remarks largely support the kind of connection we saw for loving and hating, above, in which the "acting" comes temporally prior to the "knowing":

[T2] 就如稱某人知孝、某人知弟，必是其人已曾行孝行弟，方可稱他知孝知弟，不成只是曉得說些孝弟的話，便可稱為知孝弟。又如知痛，必已自痛了方知痛，知寒，必已自寒了；知饑，必已自饑了；知行如何分得開？此便是知行的本體，不曾有私意隔斷的。

Suppose one says that someone knows filial piety or that someone knows fraternal respect. They must have already enacted filial piety and fraternal respect, and only then can they be said to know filial piety or fraternal respect. If they only understand how to say some filial or respectful words, one shouldn't straightaway say that they know filial piety or fraternal respect. Or, again, consider knowledge of pain. Only after one has been pained can one know pain. One can know cold only after one has been cold. One can know hunger only after one has been hungry. How then can knowledge and action be separated? This is just the original substance of knowledge and action, which selfish desires have not yet divided. (*IPL* 5, *QJ* 4)

Here, in describing knowledge of filiality and respect, Wang speaks of action directly. He endorses two important claims:

(Sequential KA - Filiality) If one knows filial piety, one has enacted filial piety.

(Sequential KA - Respect) If one knows fraternal respect, one has enacted fraternal respect.[13]

Henceforth, I'll abbreviate "filial piety" and "fraternal respect" and their cognates by "filiality" and "respect" and their cognates. Given the earlier discussion on which these points expand, and Wang's concluding comment on the original substance of knowledge and action, Wang clearly has genuine knowledge in mind; in context the principles should be taken to apply to genuine knowledge of filiality and genuine knowledge of respect.

[12] The attribution of this claim to Wang is suggested by a series of other passages *IPL* 226, *QJ*, 109-110 (cf. *QJ* 1292-3, *QJBB* 323), *IPL* 132, *QJ* 46-7. See below, n. 45. I address this issue in detail in "Inclinations and Mental Actions in Wang Yangming" (MS).

[13] For these principles, cf. *QJ* 1292-3.

Wang's three further examples – of pain, cold, and hunger – illustrate the idea (which can also be seen in the passage from Cheng Yi summarized above) that certain forms of knowledge require first-hand experience. If Wang intends these examples also to illustrate the connection between knowledge and action, the suggestion seems to be that being pained, being cold or being hungry should in this context be thought of as actions or as having actions as components or consequences of them. (Again, we don't need to take a stand on whether Wang would have endorsed this claim *in propria persona*; all we need here is that he is asking the reader to think this way as he presents his doctrine.) Wang's idea seems to be that one's psychological and physical responses to pain, cold and hunger can be thought of as (at least) analogous to actions, just as the psychological or physiological responses of loving and hating are actions. Allowing ourselves a little linguistic stretch, these psychological or internal physical responses might be described as "acting painedly", "acting coldly", or "acting hungrily"; analogously to the two sequential principles just presented, Wang would endorse the claim that if one knows pain, one has acted painedly, and similarly for cold and hunger.[14] A natural principle that encompasses the five examples from **[T2]** is, where "F" is to be filled in with the relevant examples:

Sequential KA If one genuinely knows Fness, one has acted Fly.[15]

Unless noted otherwise, when I present principles like this one below, I will take there to be four clear examples of relevant F: filiality, respect, compassion and loyalty. A later note (n. 73) will discuss possible further examples, but my remarks in the main text will focus on these four examples exclusively. Readers unused to schematic principles like this one may wish simply to replace "F" with "filial" throughout, and be aware that the examples are supposed to generalize also to respect, compassion and loyalty.[16]

[14] For possible connections between hunger and various actions (as more traditionally understood): see *QJ* 183 cf Ching (1972, p. 38) (猶饑者以求飽為事，飲食者，求飽之事也) A related discussion of an itchiness (that emphasizes introspection) is in *IPL* 144, *QJ* 65; cf. also the discussion of breathing and itchiness *IPL* 142, *QJ* 62.

[15] In displaying principles like this here and throughout I'm going to ask for the reader's charity in producing instances of them. Here for example "Fness" should be replaced by the appropriate abstract noun corresponding to whatever is substituted for F, e.g. "filiality" where "F" is replaced by "filial". I'll also speak of knowledge of (fraternal) respect, even though "respectfulness" and "respectfully" would be the appropriate substitutions. In large part the need for this charity has to do with imperfect correspondence between forms in English and Chinese. When I say "for relevant F", this is to be understood as limiting substitution instances of these schematic principles.

[16] Wang clearly endorses something very close to Sequential KA for filiality and respect, and it's easy to see how he might have endorsed this principle for pain, cold and hunger. But while the examples Wang draws from the *Great Learning* (loving a lovely color, hating a hateful odor) are similar in spirit to these later examples, they don't fit this model exactly. In describing those examples, Wang does speak of the action in question coming no later than the knowledge in question; in this respect they are similar in spirit. But

These "Sequential" claims are clearly supported by the text; Wang says explicitly that he endorses them. But I now want to argue that he must also have believed in a further important connection between genuine knowledge and action.[17] The "Sequential" claims allow a temporal difference between action and knowledge; they say that if one knows, one *has* acted. So if Wang only endorsed the "Sequential" principles, he would have allowed a gap between knowledge and action: he would have left it open that someone who was filial at one time, but who today is not filial any more could still genuinely know filiality. But if Wang allowed this kind of gap, then his reply to Xu Ai would not have made any sense. If Wang endorsed only the "Sequential" principles, Wang should have replied to Xu by agreeing with Xu that the kind of example he describes is possible, and elucidating why the form of "unity" he advances is compatible with the sort of cases that Xu presents.

But Wang does not reply in this way. He rejects Xu's example, and explains why the knowledge in Xu's example is not genuine. So Wang must endorse a stronger principle, which precludes this kind of example. I'll consider two styles of principle Wang might have endorsed to close this gap between knowing and acting. A first way Wang might have ruled out this kind of example would be if he accepted a principle along the following lines, restricted again to relevant F:

Dispositional KA If one genuinely knows Fness, one is disposed to act Fly when faced with a situation where that is appropriate.

So, for instance, if one genuinely knows filiality, one must be disposed to act filially in a situation where filiality is called for. On the assumption that the disposition can be present only if it has been exercised in the past, this claim would entail Sequential KA.

A different way Wang could have ruled out the kind of example that motivated Dispositional KA (where someone still genuinely knows filiality but doesn't act filially when circumstances call for it) would be to require that someone genuinely knows filiality at a time only if they are acting filially at that very moment, i.e. by endorsing:

---

the way he describes the object of knowledge in those cases is importantly different. There, he speaks of seeing the lovely color and smelling the hateful odor; he later speaks of knowing the *odor*. In neither case does he speak of knowing the *loveliness* of the color, or the *hatefulness* of the odor. But in the case of filiality and respect, he speaks of knowing the filiality and respect, not of knowing the *action* which presumably instantiates these qualities. There's nothing in the allusive text of the passage from the *Great Learning* itself which would have stopped him from speaking in a way more parallel to the key ethical examples (e.g. by saying 知好 and 知惡 or 知色之好 and 知臭之惡). Wang seems to have deliberately set up this disanalogy between these two cases. On the interpretation I'll develop, the examples from the *Great Learning* will turn out *not* to be examples of genuine knowledge; they will merely be analogous to genuine knowledge in certain respects. This aspect of my interpretation needs will be one of its most surprising features; I'll return to it at the end of section 7.

[17] Chen (1991, p. 100) believes that Wang only endorses these sequential principles, and nothing more.

KA If one genuinely knows *F*ness, one is acting *F*ly.[18]

In the case of hunger, this principle says that if one genuinely knows hunger, one is experiencing psychological responses appropriate to hunger (recall our stipulation above about the meaning of "acting hungrily"). This principle has what might seem to be the striking consequence that a perfect sage who is not acting filially at a given time would be said not to genuinely know filiality at that time.[19] One might think that whatever "genuinely know" means, a perfect sage must at all times genuinely know filiality; that is part of what it is for the sage to be perfect. But if "to genuinely know" describes an episode of knowledge, rather than the state of knowing, this line of thought misses the mark. The perfect sage does not need to be actively contemplating filiality all the time for them to be perfect; acting filially when filial action is called for is all that is required. So to the extent that episodes of knowledge require conscious events like actively contemplating their object, it would not be surprising that one would not have an episode of knowledge of filiality while not acting filially.

If Wang didn't endorse something along the lines of Dispositional KA or KA, then his reply to Xu Ai would be disingenuous. Although he could have accepted a different principle which ruled out Xu's examples, these principles seem to me simple, natural claims in the vicinity of the remarks Wang does make. So I think we can reasonably hypothesize that he accepted one of them.[20] But deciding which of them Wang endorsed turns on a hard question about the character of genuine knowledge, whether it itself is an episode of knowledge or something else, presumably at least in part a disposition to experience such episodes. In appendix A I discuss a passage (**[T16]**) where Wang defines *zhi* 知 as a kind of conscious episode. This passage makes it clear that in his more theoretical moods Wang focused on *zhi* as an episode, not as a state. Other discussions suggest that he held this view also for genuine knowledge (i.e. genuine *zhi*).[21]

[18] If one assumes that for the relevant *F*, any moment of knowing *F*ness is preceded by an earlier moment of knowing *F*ness (for example, because the interval in which one knows is open towards earlier times), then this would entail Sequential KA.

[19] The principle is consistent with a form of genuine knowledge by memory if the appropriate psychological responses can be experienced by recalling a previous episode, for example in the case of hunger if one can experience the relevant sensations by recalling being hungry. Essentially the same debate played out among classical Indian thinkers who also thought of "knowledge" as episodic (Matilal, 1986, p. 109-110).

[20] If Wang wanted to assert one of these two principles why did he assert instances of Sequential KA instead? I suspect the reason is that the position he opposed was the one that knowledge comes first and action later. As a rhetorical device, it may then have seemed more effective to assert that if anything *knowledge* comes later. In the context of arguing against the "knowledge first" position, it would have been less dramatic to say that in fact they happen at the same time. In discussing the examples from the *Great Learning*, he seems to say more clearly that the two are simultaneous, not that action comes first.

[21] Here is one kind of argument to this conclusion, which I find suggestive. In *IPL* 132 (*QJ* 46-7), Wang adamantly denies, in response to Gu Dongqiao, that knowledge comes first and action comes later. This

In the remainder of the paper I'm therefore going to assume that Wang was focused on an episode, and thus I'll discuss KA to the exclusion of Dispositional KA. But those who believe that genuine knowledge is a state and not an episode should still be able to accept the main components of my interpretation, and to alter my main claims slightly to give explanations for how Wang endorsed Dispositional KA.[22]

In what follows I will assume that a satisfactory interpretation of the unity of knowledge and action must explain how Wang endorsed KA.[23] In addition to KA, my interpretation will also explain how Wang endorsed the converse of KA:

AK If a person is acting *F*ly, they genuinely know *F*ness.

We haven't yet seen textual evidence for this claim, and I do not have space to argue here that the claim should be seen as a desideratum on any interpretation. But the idea does seem so ingrained in Wang's philosophical tradition, that there is reason to think he would have subscribed to it whether or not he said he did. Although it is today controversial whether a virtuous person must do the right thing for the right reasons, Wang and his contemporaries would naturally have thought that no person could act filially unless they were at that moment cognitively responsive in an important way to the demands of filiality. So while I won't adduce textual evidence for it in this paper, and while I won't use it as a premise in any arguments below, I do believe it is a virtue of my interpretation that it explains how Wang endorsed this claim.[24]

---

suggests that Wang would have rejected Dispositional KA, since even if one must act to acquire the disposition in the first place, later actions that exercise the disposition would come after one already possessed knowledge. It would then be true that in these cases knowledge comes first and action comes later, contrary to what Wang argues against Gu. One might attempt to bolster this line of argument using the many passages in which Wang says that the principal failing of his contemporaries (whom he sees as blindly following a certain kind of Cheng-Zhu orthodoxy) was to believe that one must first acquire knowledge, and then later act on it (see for references see n. 76). This style of argument, while interesting, strikes me as less persuasive than the one based on *IPL* 132 (*QJ* 46-7).

[22]Those who think that genuine knowledge is a state will presumably take it to be partly constituted by a disposition to experience the kind of episode which is my focus in the rest of the paper. It is easy to extend my arguments that genuine knowledge understood as an episode satisfies KA to the claim that the disposition to experience such episodes would satisfy Dispositional KA. Those who take genuine knowledge to be the relevant disposition, then, can simply use my explanations of how Wang endorsed KA to show how he endorsed Dispositional KA. So, with these minor modifications, they should be able to accept the main lines of what I say in the rest of the paper.

[23]Given the argument of n. 18, an explanation of how Wang endorsed KA would immediately give us an explanation of how he endorsed Sequential KA, so I won't say more about Sequential KA here.

[24]In Sequential KA, Dispositional KA, KA and AK, I've rendered the objects of genuine knowledge as abstract nouns. At least on the surface there are important differences between what's denoted by "know" when it's followed by a simple nominal object (for instance, knowing a person or knowing a place) and what's denoted by "know" when it's followed by items of other syntactic types: e.g., by a sentential complement (for instance, knowing that Luoyang was the capital of Tang), by a how-clause (for instance, knowing how to ride a horse), or by an infinitival expression (for instance, knowing to put the fork on the

The expression *heyi* 合一 can mean "unity", but it can also mean something more

left of a table setting). So this syntactic feature of my displayed principles is potentially significant; in fact there's an ongoing scholarly debate about whether Wang had in mind knowing-how (see Huang (2008), Ivanhoe (2000, p. 71, n. 15)) knowing-to (Huang (2017)), or something else when he spoke of the unity of knowledge and action.

I will now argue that in the passages we've considered so far Wang exclusively uses verbs of perception and knowledge followed by simple nominal complements, thereby defending the use of objectual knowledge in my key principles.

For the first three examples in **[T1]**, the point is easily established: here, the only linguistically possible construal of the three relevant expressions (見好色, 聞惡臭 and 知臭) is as describing objectual knowledge: seeing a beautiful color, smelling a bad odor, and knowing the odor. In the last three examples from **[T2]** (pain, hunger and cold), however, the situation is more complex: there are multiple linguistically possible construals of the relevant expressions. Given only the linguistic constraints of literary Chinese, Wang *could* be talking about knowing that one is hungry, knowing how to be hungry, or knowing to be hungry, as opposed to knowing hunger (and similarly for pain and cold). But the first three of these readings are obviously not plausible interpretations of the passage. The claim that one knows that one is hungry only if one has been hungry would be at best out of place here. If any instant when one is hungry is preceded by another instant when one has also been hungry, then this claim follows trivially from the "factivity of knowledge", the fact that if one knows that p, then p. But Wang doesn't say anything about the temporal structure of episodes of hunger, and the factivity of knowledge is simply not relevant here. Since Wang would have given no preparation for such an argument, this first option is really no option at all. The second and third options – that one knows how to be hungry or that one knows to be hungry only if one has been hungry – appear to be false and thus in even worse standing. Newborns and even fetuses know how to be hungry and know to be hungry at the right time, even at times when they have never been hungry before. But even if one thought Wang for some reason believed these false claims, Wang again just doesn't seem to be concerned with the kinds of issues that would be raised by them. Once again, he has done nothing to set the stage for these ideas, so the second and third options are implausible as well. While it's not clear that these exhaust linguistically possible construals of the passage, they do exhaust the reasonable options, so in these three examples too it is clear that Wang is describing objectual knowledge.

Finally, in the examples of filiality and respect in **[T2]**, again the language alone would allow that Wang describes knowing that one is filial, knowing how to be filial, or knowing to be filial (and similarly for respect). But it would be bizarre if Wang intended the descriptions of filiality and respect as discussions of knowing that, knowing how, or knowing to, but gave *five* examples uniformly of objectual knowledge to illustrate his idea. So we should understand Wang's talk of knowing filiality and knowing respect, too, as describing objectual knowledge.

(Related arguments also tell against a quite different, influential interpretation of Wang's key locutions, according to which Wang is here discussing "seeing *as* beautiful". (A prime example is Cua (1982).) Wang never discusses in this passage seeing that the color is beautiful (e.g. 見色之好) or taking the color to be beautiful (e.g. 以色為好). He uses only constructions in which the modifier either does not occur, or occurs before the noun. While it may be that this "seeing as" interpretation has been intended merely as philosophical exegesis of what Wang says, it would be striking if this is what he meant that he should not say it (there is no lack of common constructions to express the basic point, e.g. 以 X 為 Y).)

Wang's emphasis on knowledge with a simple nominal objects in these passages contrasts with the forms in Xu's opening question, where the complement of "know" (知得) is "one must be filial to one's parents" (父當孝), so that the verb there is naturally rendered by "know that", describing what we might call "propositional knowledge". In our passages, then, Wang implicitly observes a distinction between objectual knowledge – in all of the examples illustrating genuine knowledge – and the propositional knowledge Xu describes in his original example. Whether or not Wang would have articulated this difference himself, his implicit observation of this distinction, and his focus on objectual knowledge justifies rendering genuine knowledge as objectual knowledge in the principles Sequential KA, Dispositional KA and KA.

But although Wang consistently uses this construction in the passages we have seen so far, there are other passages where he speaks of his favored form of knowledge using different constructions. For instance,

like "correspondence", "correlation", or "co-occurrence". If I am right that the "unity of knowledge and action" is to be understood primarily as the conjunction of KA and AK, then perhaps "the co-occurrence of knowledge and action" would be the most natural (albeit wooden) rendering of the slogan. But the traditional translation "unity" has sometimes gone along with adherence to a stronger claim than any I've considered so far, the claim that knowledge and action are in some sense identical. Using the style of principle I've developed here, the idea would be that for relevant F:

Identity To genuinely know Fness just is to act Fly,

where "just is" is intended to denote a symmetric relation, so that Identity is equivalent to the claim that to act Fly just is to genuinely know Fness. In Appendix A I argue that Wang did not endorse this principle. In later notes, I'll consider a few principles which posit a more intimate connection between knowledge and action than that captured by KA and AK (notes 65 and 70), but none of the views I consider below will vindicate Identity, and my focus in the main text throughout will be the conjunction of KA and AK.

The goal of this section has been to put forward KA as a key desideratum on the unity of knowledge and action, and suggest AK as a natural supplement to that claim.

---

in *IPL* 118 (*QJ* 39) the favored form of knowledge is described in a way that must either be rendered "know...to..." or "know how to": "there are no children that do not know [how?] to love their parents, none who do not know [how?] to respect their brothers" (孩提之童無不知愛其親，無不知敬其兄). (This could be explained away as a fairly direct borrowing from *Mencius* 7A15.) Or again, in *IPL* 138 (*QJ* 55) (part of which is quoted and translated in n. 9 above), we have one description as propositional directly concerning ought claims, and another that is explicitly about knowing-how.

The remainder of this paper will nevertheless continue to focus exclusively on objectual knowledge. This restriction of focus helps to make the topic manageable, and is compatible with a range of attitudes to the cases where Wang uses another manner of speaking. For example one might think that in those passages Wang was making a mistake, so that we shouldn't take them seriously. Or again, one might think that there are different forms of genuine knowledge, and that this paper therefore only discusses one of them. In any case I will simply bracket these issues in what follows.

My focus on objectual knowledge may make some of what I say reminiscent of remarks of Cua, who on occasion speaks of knowledge by acquaintance, i.e. objectual knowledge (Cua, 1982, p. 7). But Cua's view of this knowledge-by-acquaintance is quite special, and different from my understanding (and in any case he doesn't adhere to the locution systematically). In Cua's view, genuine knowledge is acquaintance with the object of one's perception together with an acknowledgement that the object falls under a particular category – an acknowledgement which involves an affective response. For instance, one is acquainted with a beautiful color, and acknowledges it as beautiful in a way that involves loving it. Unfortunately, Cua doesn't provide a worked example in the case of filiality. The most natural extension of what he does say would yield the result that the relevant facts are that a person is acquainted in that case with their parents and acknowledges them as their parents. But these "relevant" facts have no place for knowledge of filiality, which was the target of our analysis. If knowing that something is beautiful involves acknowledging it as beautiful, then knowing something as filial ought to involve acknowledging it as filial. Instead it apparently involves acknowledging the object as one's parents.

Conjoined, KA and AK yield:

Unity A person is acting Fly if and only if they genuinely know Fness.

The remainder of the paper will provide a detailed analysis of Wang's moral psychology that explains how Wang endorsed Unity, and in particular shows how he gave a characterization of genuine knowledge independently of the connection between genuine knowledge and action. Sections 3 and 4 will argue that Wang held that a person's being in a particular psychological condition was both necessary and sufficient for the person to be acting virtuously. Sections 5 and 6 will then argue that Wang held that being in this psychological state is also both necessary and sufficient for having genuine knowledge. Most importantly, Section 6 identifies a new argument, "the obscuration argument", in which I believe Wang explains why he takes genuine knowledge to be privileged on distinctively epistemic or doxastic grounds. Section 7 then lays out my "introspective model", arguing that Wang took genuine knowledge to be identical to an episode of knowledge about a person's own mental life.[25] Since Wang held that genuine knowledge is a form of knowledge, and argued that it was privileged on distinctively epistemic or doxastic grounds, he was not guilty of stipulating that genuine knowledge should be understood as "whatever is unified with action". Rather, as I will argue, he advances a series of controversial and substantive theses the connection between motivational coherence, doxastic coherence and virtuous action.

3 The mind and ethical qualities

In this section I'll argue that Wang held that a person is acting filially if and and only if they have a filial mind, and more generally that for the relevant F I introduced above, a person is acting Fly if and only if they have an F mind. Elucidating this connection between the ethical quality of one's psychological condition and the ethical quality of one's action, is the first step in developing the interpretation of the unity of knowledge and action I will defend. Later sections will connect the ethical quality of one's psychological condition to genuine knowledge.

[25] Sometimes in ordinary English "introspection" describes an effortful process of directing one's attention at one's own mind, and considering its contents. I am *not* using the word in this way. Rather, I am following a standard philosophical usage according to which *any* knowledge of one's own mind counts as a form of introspection. For instance, when I later say that *liangzhi* is a capacity for knowing the rightness or wrongness of one's own mental events, I emphatically do not mean that it requires effort to exercise this capacity, or that this knowledge is conscious in any way. On the contrary, as we will see below, *liangzhi*'s knowledge of mental events is effortless, automatic and may not be conscious.

Wang taught that ethical cultivation does not require the acquisition of factual knowledge about rules for proper conduct. Many of his interlocutors expressed astonishment at this idea. Wang is often asked how a person could possibly learn how to perform proper rituals for cooling their parents in summer and warming them in winter (standard examples of expressions of filiality) without studying the details of these rituals. In one recorded response to a question of this kind, Wang retorts that if learning such detailed rituals were all that filiality required, it would take at most a day or two of study to be filial. He continues:

[T3] 若只是那些儀節求得是當，便謂至善，即如今扮戲子，扮得許多溫凊奉養的儀節是當，亦可謂之至善矣。

If [someone] who successfully seeks for their detailed ritual observances to be correct were already said [to have achieved] the highest good, it's as if an actor who successfully acts out the many detailed ritual observances of warming and cooling, serving and nurturing appropriately, could also be said [to have achieved] the highest good. (*IPL* 4, *QJ* 3-4)

Here, Wang says clearly that acting according to rules for proper conduct is not enough for one's conduct to be virtuous. His remarks strongly suggest that in his view, if a person is to act virtuously, they must be in an appropriate psychological state.

The following passage amplifies this point. Here, Wang considers two legendary historical examples, one in which the sage emperor Shun married without asking his parents' permission, and a second in which King Wu raised an army during a period of mandated mourning. Wang clearly holds that although these behaviors violated the standard rules of filiality and loyalty, respectively, nevertheless the first was filial and the second was loyal. But he writes:

[T4] 使舜之心而非誠於為無後，武之心而非誠於為救民，則其不告而娶與不葬而興師，乃不孝不忠之大者.

Suppose Shun's mind had not been sincere with regard to not having descendants, or Wu's mind had not been sincere with regard to saving the people. Then Shun's marrying without telling [his parents] or Wu's raising an army without mourning would have been enormously unfilial and disloyal. (*IPL* 139, *QJ* 57)

The word I've followed tradition in translating as "sincere" here and throughout might be better rendered by "wholehearted": as we will see in more detail below, in saying that Shun and Wu's minds were sincere, Wang seems to mean that they did not experience psychological conflict as they performed the relevant actions.

This passage supports attributing to Wang the view that an appropriate state of mind is required for a person to act virtuously. It suggests a more specific version of the idea than we saw in **[T3]**, namely, that if an action is filial or loyal, the person who performs it must in particular have a mind which is sincere (or wholehearted) with regard to that action. For Wang says explicitly that even if Shun or Wu had performed exactly the same physical actions in exactly the same mind-external situations, these actions would have been unfilial and disloyal if Shun and Wu had not had minds which were sincere with respect to these actions.

In addition to the claim that acting virtuously requires a person to be in an appropriate state of mind, the passage also provides some evidence that Wang held the more controversial claim that being in an appropriate state of mind is sufficient to guarantee that a person will act virtuously. Shun's and Wu's actions were at the very extremes of unfiliality and disloyalty; one can hardly imagine actions that would be more paradigmatic violations of the requirements of these virtues. But Wang condones them as filial and loyal, and the basis for his positive appraisal is that both Shun and Wu had minds which were sincere. So the passage suggests that Wang held that if any action – no matter how apparently vicious – is performed with a mind which is sincere, then the action will be virtuous. (This doesn't mean anything goes; as we'll see later he believed there were limits to what actions could be performed with a sincere mind.) [26]

[26] A number of authors, most notably Angle & Tiwald (2017, p. 60-1), but see also Nivison (1996a, p. 340), have taken Wang to be espousing a kind of situationism or particularism in **[T3]**, **[T4]** and related passages. On one reasonable regimentation of the terms, *generalism* is the claim that a person is acting virtuously only if they are considering a general principle of ethical conduct which governs their action; *particularism* is the negation of generalism. In the lead-up to **[T4]** Wang does emphasize that Shun and Wu did not have examples to base their conduct on, and thus had to follow their *liangzhi*. But while one might conclude from this that Shun and Wu did not follow general ethical principles at all as they acted, a different moral to draw is that they did not have to *learn* these ethical principles, since they already knew them innately in virtue of their *liangzhi*, a conscience-like aspect of the mind I'll discuss in much more detail below. Indeed, this point seems closer to the surface of Wang's discussion, since in the passage immediately preceding his mention of these examples he analogizes the way one should use *liangzhi* to judge good and evil, to the way one should use the compass and the ruler to measure roundness and length. The parallel suggests not that there are no standards for virtuous conduct, but that the appropriate tool for ensuring that one accords with the standards there are is one's own *liangzhi* (and not some kind of acquired knowledge of detailed ritual observances).

Angle and Tiwald see evidence for particularism also in the first sentence of Wang's "four sentence teaching" (*IPL* 315 *QJ* 133, "the original substance *ben ti* of the mind is without good and evil" see above n. 7 for the translation "original substance"). They take this passage to mean that "external, rigid standards of good and bad have no place in the heartmind's inherent reality [their translation of *ben ti*, which I've translated 'original substance']" (61). But this passage seems to me to make a very different point. In the surrounding texts, Wang never explicitly mentions rigid, external standards or general ethical principles. A more natural interpretation seems to me to take the text at face value: Wang's point is (as he says) that good and evil do not apply to the original substance of the mind, rather (as he goes on to say in the second sentence of the four sentence teaching), they apply primarily to inclinations ("when inclinations arise, there is good and evil"). Inclinations arise from but are not identical to the original substance of the mind. On

A series of further passages suggest that Wang held not only that a person acts virtuously if and only if their mind is in an appropriate state, but also that if a person's action is virtuous, it is virtuous *because* they are in an appropriate psychological state. In a somewhat picturesque idiom, we might say that if a person's action is virtuous, the state of the person's mind is what makes it so. While this claim about explanation or "making it so" will not be essential to my interpretation of the unity of knowledge and action, it merits our attention in this context for two reasons. First, these passages provide some of the clearest evidence for the claim that a certain condition of a person's mind is sufficient for their physical behavior to be virtuous, and this claim *will* be essential to my interpretation below. Second, the fact that Wang held that the condition of one's mind explains the virtuousness of one's actions will help us to understand how Wang could endorse this claim – how he could have thought that a particular condition of one's mind could be sufficient for one's actions to be virtuous. For it might seem that, no matter how good one's intentions, one might still act wrongly because of the vicissitudes of one's circumstances. But Wang argues on conceptual grounds that this is not possible, since what it is for an action to be virtuous is for it to be performed in a certain state of mind. The actual physical performances which result are not relevant to whether they are virtuous or not.

In the passages we'll now examine, Wang often uses the term "*li*", a central term in what is now called "neo-Confucian" metaphysics, that has been variously translated "principle", "pattern" and "coherence".[27] The concept of *li* is one of the most difficult in Wang's philosophical tradition: just about any interesting claims related to it will be highly contentious. But even so it will be worth seeing how the present interpretation could help to make sense of these difficult passages, and how they might support the claim that Wang held that the mind explains the virtuousness of virtuous actions.

[T5] 夫求理於事事物物者，如求孝之理於其親之謂也。求孝之理於其親，則孝之

a more straightforward reading, then, the text makes the metaphysical point that the appropriate bearers of these qualities are certain mental events; it is not concerned with the applicability or inapplicability of general rules.

Sometimes (although not, I believe, by the authors cited here) the term "particularism" is associated with the view that there are no general principles governing what actions are virtuous or what makes an action virtuous. But Wang surely was not a particularist in this sense: he held that there *are* general rules relating the state of one's mind to the ethical character of the action performed (for instance, good actions are those performed with a sincere mind), and (as I suggest below) that the condition of one's mind is what makes an action virtuous if it is. Indeed, as I will be arguing, Wang seems to endorse a position that we might call *intentionalism*: the view that the ethical character of a person's physical actions is entirely determined by the person's psychology at the time of acting.

[27] "Principle" is perhaps the standard translation, seen for instance in Chan (1963). "Pattern" is used throughout Angle & Tiwald (2017). "Coherence" is due to Peterson (1986).

理其果在於吾之心邪？抑果在於親之身邪？假而果在於親之身，則親沒之後，吾心遂無孝之理歟？見孺子之入井，必有惻隱之理，是惻隱之理果在於孺子之身歟？抑在於吾心之良知歟？其或不可以從之於井歟？其或可以手而援之歟？是皆所謂理也，是果在於孺子之身歟？抑果出於吾心之良知歟？以是例之，萬事萬物之理，莫不皆然。

Looking for *li* in each individual thing and affair is like looking for the *li* of filiality in your parents. If you look for the *li* of filiality in your parents, then is the *li* of filiality really in your mind? Or is it really in the person of your parents?[28] If it is really in the person of your parents, then immediately after they have passed away will your mind have no *li* of filiality? If you see a child fall into a well, there must be a *li* of compassion. Is this *li* of compassion really in the person of the child? Or is it in the *liangzhi* of your mind? Is it impossible for you to follow it in to the well? Or is it possible that you can pull it [out] with your hand? All of this is what is called *li*. Is it really in the person of the child? Or does it rather proceed from the *liangzhi* of your mind? Taking this as a model, [we can see that] the *li* of all things and affairs are like this. (*IPL* 135, *QJ* 50-1 cf. *IPL* 3, *QJ* 2-3)

I'll come back to Wang's notion of *liangzhi* in more detail in the next section. For now the passage should be legible treating it as a black box.

There was a longstanding idea – emphasized by Zhu Xi and others – that *li* is what makes things the way they are (所以然者). For instance, the *li* of water is what makes water cold, the *li* of fire is what makes fire hot, the *li* of a boat is what makes a boat only able to go on water; the *li* of carts is what makes a cart only able to go on land.[29] If Wang was drawing on this conception of *li* then in his repeated, strenuous arguments that *li* is internal to the mind he may have been arguing at least in part for the claim that if a person's action is filial, it is filial because of the condition of the person's mind.[30]

If Wang has this aspect of *li* in mind, much of the passage makes good sense. In the first sentence he connects *li* in general to the specific *li* of filiality. Wang seems to assume that if a person's actions at a time are filial, what makes them filial must exist at that time. He then argues that the *li* of filiality cannot be in the parents. One of the most

[28]The expression 在於 translated here as "is in" can mean something weaker like "depend on". But later in the passage Wang uses 出於 "emanate from" which strongly suggests he is thinking of the locative meaning of the expression. (In closely related passages, e.g. **[T6]** quoted next, he also seems to use the expression in a locative sense.) Even if this were wrong, in the present passage Wang quite clearly takes the alternatives to be exclusive of one another, so even if he is using the word to mean "depend", in context it means "completely depend".

[29]For water and fire see passage cited on Graham (1958, p. 75).For boats and carts, see *ZZYL*, 61 (問：「理是人物同得於天者。如物之無情者，亦有理否？」曰：「固是有理，如舟只可行之於水，車只可行之於陸。」).

[30]Song thinkers also described *li* as determining moral qualities; it is "that by which things are loyal, trustworthy, filial and respectful" *ZZYL*, 124 (所以為忠信孝弟者) cf. *ZZYL* 585.

important manifestations of filiality was obeying ritual observances of mourning, so it must be that a person's actions can be filial after their parents have died. But then what makes a person's actions filial cannot be in the person's parents. If it were, it would cease to exist along with the person's parents, yielding the absurd result that actions taken after a person's parents' death could no longer be filial.

Wang then turns to a canonical example from Mencius, who famously argued that people's nature is intrinsically good by claiming that anyone who sees a child on the verge of falling into a well would naturally feel compassion for the child. Wang begins by claiming that when one sees the child there must be the *li* of compassion. His idea again seems to be that when one sees the child, one will feel compassion, so that there must be something which makes this feeling of compassion (and any actions which issue from it) compassionate, that is, the *li* of compassion.

Wang asks where the *li* of compassion is located in this scenario, and argues that this *li* cannot be in the child. His argument is laconic, but it seems to turn on the claim that the child does not determine the feasible set of actions for the bystander. He starts by observing that whether it is possible for one to save the child or not is called *li*. His idea seems to be that what it is to show compassion for the child may differ depending on one's physical relationship to the child. Perhaps one is next to the well, and one must do one's best to save the child, or perhaps one is at a great distance, and there is nothing to be done, so that all compassion requires is an appropriate affective response. But what physical (and mental) actions are appropriate for a bystander is not something determined by the child, since the child does not determine (for instance) whether it is possible for the bystander to save the child. So, the *li* of compassion cannot be the child or in the child. Wang jumps to the conclusion that what makes the action compassionate must be something in the mind, and indeed something in the *liangzhi* of the mind. This is a significant leap: after all, even if the *child* doesn't determine whether the action is compassionate, couldn't some set of mind-external factors nevertheless determine it?[31] But Wang's goal is not to give a deductive argument for his conclusion here. He seems rather to appeal to the reader's preference for a simpler theory to motivate his idea, and to invite us to reason as follows: which actions are compassionate can vary markedly across different physical circumstances, but these different compassionate actions have in common the condition of the person's mind, so the most natural conclusion to draw is that what makes the actions compassionate is the condition of the person's mind. Of course – we could imagine Wang saying – the pedantic reader might try to stake out a laundry list of external conditions and which actions are compassionate in them,

[31] So too, Chen (1991, p. 27).

claiming that it is these heterogeneous features of the external conditions that make the actions virtuous. But isn't it simpler (he might continue) to appeal directly to whether the person felt unstinting compassion when they acted as the determinant of the quality of their actions? If Wang's reasoning were of this more abductive kind, he is not "skipping" a deductive step, but simply inviting the reader to recognize the striking commonality in the examples. But whether or not Wang endorsed this imagined train of thought, his conclusion is clear: what makes the action compassionate (the *li* of compassion) is in the mind. And he says we may take this example as a model, suggesting that he would have endorsed the claim that quite generally the ethical quality of a person's actions is determined by the quality of their mind when they were acting.[32]

In **[T5]** Wang presents the general idea that *li* are in the mind, but doesn't expand on what feature of the mind makes actions virtuous if they are. The next passage offers further remarks along these lines:

[T6] 夫物理不外於吾心，外吾心而求物理，無物理矣。。。。。故有孝親之心，即有孝之理，無孝親之心，即無孝之理矣。有忠君之心，即有忠之理，無忠君之心，即無忠之理矣。理豈外於吾心邪？。。。。。

心，一而已。以其全體側怛而言謂之仁，以其得宜而言謂之義，以其條理而言謂之理；不可外心以求仁，不可外心以求義，獨可外心以求理乎？外心以求理，此知行之所以二也。求理於吾心，此聖門知行合一之教，吾子又何疑乎？

The *li* of things are not external to your mind. If you look for the *li* of things outside your mind, there will be no *li* of things...Thus, if there is a mind which is filial to your parents, then there is the *li* of filiality. If there is not a mind which is filial to your parents, there is no *li* of filiality. If there is a mind which is loyal to your lord, then there is the *li* of loyalty. If there is not a mind which is loyal to your lord, there is no *li* of loyalty. Are *li* external to the mind?...

The mind is one. In terms of its complete empathy, it is called humane. In terms of its achieving what is appropriate, it is called righteous. In terms of its orderly pattern (*li*), it is called *li*. You cannot look for humaneness outside the mind. You cannot look for righteousness outside the mind. Should only look for *li* outside the mind? This is what it is to take knowledge and action to be two. To look for *li* in your mind is the sages' teaching of the unity of knowledge and action. How can you still doubt it? (*IPL* 133, *QJ* 48)

In the first part of this passage, Wang emphasizes the connections between the *li* of filiality and the mind which is filial toward one's parents (孝親之心), and between

[32] A striking passage from his second letter to Wang Chunfu, where Wang says that there are no *li* outside the mind (心外無理 *QJ* 175, translated in Ching (1972, p. 29-30)) further supports this claim. His point in context seems to be that there are no determinants of ethical qualities outside of the condition of the mind.

the *li* of loyalty and the mind which is loyal to one's ruler (忠君之心).[33] Wang says explicitly that the *li* of filiality is present at a time if and only if there is the mind which is filial towards one's parents at that time. If we understand Wang's remarks about *li* as intended to describe that which makes something what it is, the passage suggests directly that actions which are filial or loyal are filial or loyal because the person who performs them has a mind which is filial or loyal.

[T3] and **[T4]** provide strong evidence that Wang held that acting virtuously requires one's mind to be in a certain condition. Our two passages about *li*, **[T5]** and **[T6]**, provide further support for this idea. If an action is filial, there must be the *li* of filiality, i.e. what makes it filial. And this *li* of filiality is present if and only if the person who acts has a mind which is filial. So if a person's action is filial, the person has a mind which is filial.

[T4] also provided some – albeit more tenuous – evidence for the claim that Wang held that a certain condition of the mind was sufficient for a person's action to be virtuous. The present passage importantly augments the case for this second claim. If Wang thought of the *li* of filiality as that which makes an action filial or explains the filiality of a filial action, it is extremely likely that he would have held that if this *li* of filiality is present in a person, then the person's action is filial. Since he clearly says that the *li* of filiality is present if and only if a person has a mind which is filial toward their parents, he would have held that a mind which is filial is on its own sufficient for a person's

[33] The key expression "the mind which …" (… 之心) has two quite different uses in Wang's writings. In some of them, for example, "the mind which judges right and wrong" or "the mind which approves and disapproves" (是非之心), an expression which occurs in the *Mencius* (most notably in 2A6), Wang seems to use the expression to describe the mental capacity for *X*ing. Someone may have the mind which approves and disapproves even if they are not now approving or disapproving of anything. Other uses, however, mean a mental state or episode of *X*ing. This interpretation is essentially mandated by a passage in *IPL* 105 (*QJ* 35), where Wang says "if the mind that devotes itself to substantial things becomes weightier, then the mind that devotes itself to reputation becomes lighter; if [one's mind] is entirely the mind that is devoted to substantial things, then there will not be any mind that is devoted to reputation" (務實之心重一分，則務名之心輕一分；全是務實之心，即全無務名之心). Here Wang describes the waxing and waning of a particular mental state (for example, a standing desire) rather than the waxing and waning of a general capacity for doing these things: a person does not lose their capacity for devoting themselves to reputation simply because they no longer do so.

The most natural way of understanding these locutions in the passage above is as describing a mental state or episode, as opposed to a capacity. For Wang talks about there being and there not being this mind; if we take him to be talking about the faculty then he would be contemplating a possibility in which humans are deprived of an intrinsically human capacity. But such an outlandish idea is far from the train of thought in the passage. In the main text I will therefore assume that we adopt this latter interpretation, on which "the mind which..." is understood as a state of the mind.

The notion of mental states or mental events common in contemporary philosophy does not line up with events which Wang would have seen as produced by the *xin* 心 (which I've translated as the mind). I have been and will continue to be careful to speak in a way that fits both usages but a reader who thinks in terms of the modern category won't miss anything particularly important, and I won't comment explicitly on the issue further.

actions to be filial: if a person has a mind which is filial, then any actions they perform are filial. Earlier, we saw some evidence that Wang would have held that a person is acting virtuously if and only if they have a mind which is sincere with regard to their action. The present passage provides stronger evidence that Wang would have held a related claim: that a person has a mind which is filial (or: loyal) if and only if they are acting filially (or: loyally).[34] More generally, the passage suggests, for relevant F,

Mind Action A person has a mind which is F if and only if they are acting Fly.

The passages in this section demonstrate that Wang held that the qualities of a person's mind are intimately related the qualities of a person's action. In **[T4]** Wang describes the mind which is sincere; in **[T6]** he describes the mind which is filial and the mind which is loyal. In the next section I will examine more closely what Wang meant by a mind which is sincere, filial and loyal. That examination will lead us to a crucial further connection between the mind and action, which will be the first thesis of my introspective model of the unity of knowledge and action.

4 The sincere mind and sincere inclinations

Wang's listeners would have immediately connected his discussion of the sincere mind to the discussion of "making the inclinations sincere" 誠意 in the *Great Learning*, a canonical text known by rote by literate scholars of Wang's day. The passage from which the examples of loving lovely colors and hating hateful odors was drawn occurs in the discussion of making the inclinations sincere. In our attempt to understand what Wang means by a sincere mind, it is natural then to turn to his discussion of *yi*, inclinations.

In the philosophical tradition in which Wang wrote, one can find at least two promi-

[34] In many other passages Wang also ascribes filiality or loyalty more directly to the mind, often using the expression "the mind which …" in a way similar to the way he uses it here. In *IPL* 38 (*QJ* 17) Wang says "the mind's arousal (*fa* 發) when it encounters one's father, is said to be filial, and when it encounters one's ruler, is said to be loyal (*zhong* 忠)" (心之發也，遇父便謂之孝，遇君便謂之忠). In a letter to Zhu Yangbo, Wang uses language that closely parallels that of the passage in *IPL* 38 discussed above to describe *li* instead of the mind *xin*: "*Li* is the organizing *li* (條理) of the mind: if this *li* is aroused in regard to one's parents then it is filial; if it is aroused in regard to one's ruler then it is loyal; if it is aroused in regard to one's friend's, then it is trustworthy" (理也者，心之條理也。是理也，發之於親則為孝，發之於君則為忠，發之於朋友則為信, *QJ*, 308). The connection between *li* and the mind in exactly this context further supports the idea in the main text. In *IPL* 3 (*QJ* 3)) he speaks of "exhausting the filiality of the mind" (盡此心之孝), "the mind which is sincere with regard to being filial to one's parent's" (誠於孝親之心), and (most clearly) "the mind which is sincerely filial" (誠孝之心). (Cf. 孝親之心, *QJ* 1295.)

nent, distinct uses of the character *yi* 意.[35] In one of them, *yi* indicates the generic activity of the mind. Translators who have emphasized this feature of the expression have rendered it as "thought"; interpreters render it as "idea" or "conscious mental activity" (contemporary 意念, 意識).[36] In a second use, *yi* are more closely connected to actions. Translators who have focused on this use of the expression have typically rendered it as "intention" (contemporary 意向).[37] But I think in this use the word is better rendered as "inclination" than "intention".[38] Here is one reason why. Normally if one intends to do something, one will do it when the opportunity presents itself. One cannot under normatively typical conditions simultaneously intend to open a door and stand in front of it unimpeded without opening it. But the passage below suggests that Wang held that one *can* have an *yi* to do something, but not begin to do it, even when the opportunity presents itself:

[T7] 蓋鄙人之見，則謂意欲溫凊，意欲奉養者，所謂意也，而未可謂之誠意。必實行其溫凊奉養之意，務求自慊而無自欺，然後謂之誠意。

In my humble opinion, what is called the motivating desire to warm or cool, or the motivating desire to serve and nurture, is what I call an inclination. But you can't yet call it a sincere inclination. One must enact the inclination to warm, cool, serve or nurture, and devote yourself to seeking to be naturally content and without self-deception, and only then can it be called a sincere inclination. (*IPL* 138, *QJ* 55)[39]

[35] Qian (錢明) (2005) provides an extremely useful survey of these uses. Shun (2010, p. 179) presents a view of Zhu Xi's use of the expression that is broadly similar to what I attribute to the tradition here.

[36] e.g. Shun (2011), Chen (1991).

[37] e.g. Ching (1976), Angle & Tiwald (2017). Chen (1991) renders this use 意欲, and sometimes 意向. The standard translation of Chan (1963), "the will", as well as the "volition" of Mou (1973) also emphasize this aspect of the word. But in many constructions these expressions denote capacities, whereas *yi* is typically a mental event.

Shun (2010, p. 180) suggests that the translation "thought" (designed for the first, generic use above) can be made to fit with this second use of the word, if we restrict attention to locutions such as "I thought to hit him". Two concerns have led me not to adopt this ingenious proposal: first, since the "'think to" uses are comparatively unusual in English it is quite difficult to focus on them; second, the translation makes it hard to appreciate the contrast between *yi*, which very often motivate one to act, and *si* (思), which in the most common uses, do not (see below n. 54).

[38] Others have recognized other differences before: see most notably Qian (錢明) (2005), but also Shun (2010, p. 180).

[39] I interpret *yiyu* 意欲 as a compound here, and translate it "motivating desire". The first character of this compound is *yi* 意, the word I translate as "inclination" throughout. Chan (1963) translates the expression as a noun followed by a verb; in my idiom, "inclination which desires". In other passages Wang uncontroversially uses the expression as a compound, so I have translated it as one here as well (e.g. in "Another reply to Lin Zezhi": 今方欲與朋友說日用之間，常切點檢氣習偏處、意欲萌處，與平日所講相似與不相似，就此痛著工夫，庶幾有益。陸子壽兄弟，近日議論，卻肯向講學上理會, where it however occurs as a noun, not a verb). But the question whether it is a compound or not in this passage does not affect any important claims I make. What is important is that 意欲 had not at this point taken on its contemporary meaning of "intention", but this is clear from the passage itself.

It would be odd to deny that one's *yi* to help one's parents is sincere (or: wholehearted) only because one has not had the opportunity to cool them (perhaps it is winter when cooling is unnecessary). Wang's thought here seems to be instead that one's inclination to cool fails to be sincere if, when the opportunity presents itself, one still fails to act. So this passage suggests that *yi* are less tightly tied to action than intentions, and in this respect more like inclinations.[40] Some further differences between *yi* and intentions are discussed in a note.[41]

I will focus on the second, action-connected use of this term in what follows, and will therefore use "inclination" as my translation of *yi* throughout. I am interested in the claim that even when Wang used the word for the generic activity of the mind, he used it to mean (something like) "inclination".[42] But my arguments will not rely on this claim; I will leave it open that Wang sometimes used the term as a generic term for mental activity without intending to claim that all mental events are inclinations (as opposed to, say, thoughts).

There is one important difference between *yi* and both inclinations and intentions.

[40] A different interpretation is that *yi* are intentions, but that Wang means to describe here the difference between an intention and an episode of trying. The arguments in the next note tell against this alternative.

[41] Here are three:

1. It is rationally incoherent to intend to perform one action *and* intend to perform another which one knows to be incompatible with the first, but it is of course completely coherent to have some inclination to perform each of two actions one knows to be incompatible. *Yi* are like inclinations in this respect: it is fairly clear that one can coherently have competing *yi* for actions which one knows could not be performed together. (This strikes me as obvious, but I haven't been able to find direct evidence in Wang's writings themselves. To the extent that inclinations and concerns are similar in Wang's philosophical idiom (on "concerns", see below n. 54), the point follows from his remarks about a person's having three conflicting concerns on *QJ* 1003.)
2. Wang often speaks of *yi* as aroused or moving (e.g. **[T16]**, *IPL* 174, *QJ* 86-7). It is much more natural to speak of inclinations as aroused and moving than to speak of intentions in this way.
3. Most tentatively, we often don't think of ourselves as identifying with our inclinations. Inclinations are more like desires; they happen to you ("the urge came upon me"). Intentions can of course be formed spontaneously, we do not think of them as coming upon us as it were from outside: our intentions are ours, however they are formed. (Qian (錢明) (2005) also makes this point. Shun (2010) describes *yi* as "more conscious and deliberate" than desires 欲, but this is compatible with the above contrast between *yi* and intentions.)

The question of whether colloquial use of *yi* tracks "thought" or "intention" better than "inclination" is beyond the scope of the present investigation. There are some uses of the word in Wang that do not fit well with "inclination". For instance, in the opening sections of a letter in *IPL* 144 (*QJ* 64), a colloquial use of *yi* is closer to "intention" than "inclination" (備道道通懇切為道之意); it wouldn't be natural to speak of an inclination to "pursue the way" (為道), or a "keen" (懇切) inclination. A quite different use of the word – which is not obviously relevant to the dispute over "intention" and "inclination" – is as "aim". See for instance later in the same letter 在兩生則亦庶幾無負其遠來之意矣 (*IPL* 144, *QJ* 64).

[42] Qian (錢明) (2005) seems close to this idea, when he takes the somewhat unusual view that in its generic use the expression *yi* means "desire" (慾望). I will continue to explore this possibility on occasion in what follows, and hope to discuss it further in future work.

If someone intends to celebrate the lunar new year with their family, they intend to do this when they fall into a dreamless sleep, and quite generally whether they are considering the question of who they will spend the new year with or not. Similarly for someone who has not yet decided who they will celebrate with, and has only an inclination to spend the new year with their family, although they recognize that there are also countervailing considerations. Intentions and inclinations are fairly stable: they may last for a long time, and they persist regardless of whether there are conscious episodes associated with them. But there are also mental episodes associated with both intentions or inclinations. The intention to go home for the holiday may manifest itself in a bout of worry about the need to buy one's tickets. Similarly, the inclination to see one's family might manifest itself in a sensation of longing for one's family members. As this discussion illustrates, there is an important distinction between intentions and inclinations, and the sensations or conscious episodes associated with them. But when Wang speaks of *yi* he often (following tradition) speaks of *yi* arising (*fa*, 發), or moving (*dong* 動). These usages strongly suggest that, when Wang discusses *yi*, he is primarily focused on episodes, presumably sensations associated with underlying inclinations, and not the (standing) inclinations themselves. I have already made a similar point above, in connection to the contrast between knowledge and episodes of knowledge: in the case of knowledge, too, Wang seems to be focused more on mental episodes than he is focused on standing states. The fact that Wang discusses primarily inclination-episodes by contrast to the standing inclinations may even be taken to support the view that he is also focused on knowledge-episodes, since it may seem more generally to support the idea that he focused on mental episodes rather than mental states. While I am attracted to this general idea, I will not have space to pursue it further below, and have only wanted to emphasize here the point about *yi*.

We found in **[T7]** a negative point about the lack of connection between ordinary *yi* and action. But the passage also makes a positive point about the connection between *sincere* (or "wholehearted") *yi* and actions. Wang says that an inclination to warm or cool is sincere only if one does in fact warm or cool. The tight connection Wang postulates between action and sincere *yi* makes it natural to think of *sincere yi* as more strongly connected to action even than intentions; they are *guaranteed* to result in a relevant bodily action. Since (as I have argued) Wang thinks of inclinations themselves as mental episodes, it is natural to think of sincere inclinations as episodes of trying – mental events which are guaranteed to result in whatever bodily action the person can effect in the circumstances.

In the previous section, I suggested that Wang held that a sincere mind is sufficient

for whatever bodily action issues from it to be virtuous (regardless of what that action is). If we assume, as it is quite natural to do, that Wang held that one has a sincere inclination if and only if one has a sincere mind, then Wang would be committed to holding that if a person has a sincere inclination they will act virtuously. If one thinks of a sincere inclination as an episode of trying (as I have suggested we should), Wang's thought is the familiar (although controversial) idea that the adventitious physical circumstances of one's actions should not determine whether one's actions are virtuous. All that matters to whether the action is virtuous or not is one's state of mind, for instance, whether one tried one's best. We have seen a strong case in the last two sections that Wang held that if one's mind or one's inclinations are sincere, one will be virtuous. He seems to think that the consequences of the action one undertakes with such a mind do not add or detract from the virtuousness of the action. In Wang's view, it is the thought – or rather the *yi* – that counts.

Just as I have argued that Wang and his readers would have connected a sincere mind to sincere inclinations, one might hope to draw a similar connection between a filial mind and filial inclinations (or, sincere filial inclinations). But I don't know of a passage where Wang says explicitly that an inclination is itself filial or loyal.[43] He seems to prefer to speak, as here, of inclinations to perform actions, where those actions happen to be filial. And as we have seen, not every such inclination will issue in the relevant action; only those which are sincere are guaranteed to. In other words, Wang's remarks suggest not the direct analogue of Mind Action, but instead (again for our relevant F):

Inclination Action A person is acting Fly if and only if they have an inclination to perform an F action and that inclination is sincere.

It is important in assessing this principle to bear in mind that Wang thought of "inclinations" as mental episodes. A person might have a wholehearted inclination to do something in the distant future, but this does not guarantee that they are already acting at the present time. This style of counterexample does not affect Wang's claim, however, since his claim concerns the mental episode associated with this wholehearted inclination, an episode which occurs in the moment when action is called for. The episode

[43] He does often speak of the ethical qualities of inclinations, e.g. in the key passages *IPL* 101 (*QJ* 33-4) and *IPL* 315 (*QJ* 133). Moreover, the passages in which Wang discusses the loyalty and filiality of the mind (see n. 34) are often tightly connected to inclinations. In some of those passages, he speaks of sincerity, which as we've seen is thought of as a key property of inclinations. In others of them, he speaks of "arousal", and in many other places, Wang says that inclinations are the arousal of the mind (e.g. *IPL* 6, *QJ* 6-7)). So it may be that Wang would have found such a description congenial, even though he himself does not seem to have used it.

associated with this inclination is sincere (wholehearted) only if one is acting at that moment.

Inclination Action will be one of two key claims of my introspective model of the unity of knowledge and action.[44] One might think of it as my interpretation of the "action" half of the doctrine. In the next three sections I will turn to the "knowledge" half.[45]

[44] A further argument for attributing this claim to Wang is that it helps to bring together a number of doctrines we have seen him endorsing. Here is one argument of this kind. It is natural to think – and fairly strongly supported by reflection on **[T4]** – that Wang held (*) that a person has a filial mind or a loyal mind only if they have a sincere mind. This claim can be derived using Mind Action, Inclination Action, and a natural principle connecting the sincere mind to sincere inclinations. By Mind Action, a person has a filial mind if and only if they are acting filially. By Inclination Action, a person is acting filially if and only if they have an inclination to perform a filial action, and that inclination is sincere. So, a person has a filial mind if and only if they have an inclination to perform a filial action and that inclination is sincere. Provided that a person has an inclination which is sincere if and only if they have a sincere mind, (*) would follow. The fact that Inclination Action (together with other natural assumptions) allows us to derive (*), which Wang most likely held, is some evidence in favor of it.

Throughout this section I've presented my arguments for Inclination Action as building on or even dependent on the attribution of Mind Action to Wang. But all that will matter in what follows is whether Wang accepted Inclination Action. One might deny that Wang accepted Mind Action while holding that he accepted Inclination Action. If one did, one should still not see the discussion of the previous section as irrelevant: the passages discussed there provide circumstantial evidence for the connections Wang drew between the condition of the mind and the ethical qualities of one's actions. That circumstantial evidence is also a kind of evidence for Inclination Action itself, which relies on a tight connection between the qualities of a person's mental events, and the qualities of their actions.

[45] I have argued that Wang held that ordinary bodily actions are virtuous, if they are, because the mind of the person who performed them is virtuous. Wang may also have held that mental episodes themselves are actions. We have already seen hints of this in **[T1]** and **[T2]**. But there are clearer texts where Wang suggests more directly that mental events are actions (*IPL* 226, *QJ*, 109-110 (cf. *QJ* 1292-3, *QJBB* 323); *IPL* 132, *QJ* 46-7). From these and other passages, there is a case to be made that Wang would have endorsed:

Mental Action Inclinations, concerns, desires, thoughts, loving, hating, and emotions are all actions

This principle is plausible only if one understands inclinations, concerns, desires, thoughts, loving, hating and emotions as mental episodes rather than mental states. But as I have said, Wang's discussions of these aspects of the mind strongly suggests that he was most focused on mental episodes, as opposed to mental states. If Wang accepted Mental Action, it would have opened a range of theses connecting knowledge to action more intimately than the ones in the main text. I discuss these possible further commitments in "Inclinations and Mental Actions in Wang Yangming" (MS), but here I will continue to focus on KA and AK.

One might worry that if Wang endorsed Mental Action, then he could not have endorsed Inclination Action, since it might be possible that one could for example have a concern to act filially – and so act filially – without having an inclination to act filially. In response to this worry I would like to make two points. First, even if Wang believed that concerns are actions, he need not have believed that filial concerns are filial actions. Someone who believes that murderers are people need not believe that good murderers are good people. Second, there is a strong case to be made that Wang held that concerns, desires, thoughts, loving, hating and emotions are inclinations, or at least that they arise only if inclinations do. On this natural hypothesis, even if a filial concern counts as a filial action, it would be taken to arise with a filial inclination, and so pose no counterexample to Inclination Action. In what follows I won't consider this line of attack on Inclination Action; in "Inclinations and Mental Actions in Wang Yangming" I discuss both of these responses at greater length.

5 *Liangzhi* and knowledge of ethical qualities

Our target for the next three sections will be genuine knowledge of ethical qualities such as filiality and respect. In this section I begin developing an interpretation of genuine knowledge by examining Wang's discussions of (ordinary) knowledge of ethical qualities. I will argue that Wang held that *liangzhi* 良知 automatically knows ethical qualities – and especially the rightness and wrongness – of mental events. At the end of this section I argue that *liangzhi*'s knowledge does not invariably amount to genuine knowledge; the following section will describe what further conditions are required for this knowledge to be genuine knowledge.

The word *liangzhi*, which is made up of two characters, "pure" *liang* and "knowledge" *zhi*, occurs first in *Mencius* 7A.15, where its innateness is emphasized. The expression was used on its own comparatively rarely by Wang's predecessors; it is typically used in a pair with "good ability", *liangneng* 良能, as Mencius used it.[46] The notion was always closely connected to *zhi* because of its etymology. And, as we will see, Wang retains that connection.

To this point we have only discussed uses of *zhi* as a noun in which it stands to verbal uses of *zhi* in a way roughly similar to the way in which the English "knowledge" stands to "know": if one knows (*zhi*s) something, one has knowledge (*zhi*) of it. But as a noun *zhi* can also be used to refer to the capacity for *zhi*-ing,[47] or to the activity of that capacity's being exercised (the activity of *zhi*-ing, analogous to the exercise of a capacity of sight in the activity of seeing). *Liangzhi* can be used in all of these ways. But by far its most common use is as a noun in the first way just discussed: indicating a capacity.[48] In the passage we'll consider next, for example, Wang explicitly describes it as a capacity.[49]

This fact about the expression *liangzhi* 良知 marks an important contrast between it

[46] For a compact discussion of Wang's predecessors, see Peng (彭國翔) (2003, 30-37).

[47] A canonical citation for this linguistic point is the (much earlier) Mohist Canons A3 and A22. Interestingly, in A23, it is said that the capacity is inactive in dreamless sleep. For discussion, see Fraser (2011, p. 134).

[48] There's a parallel to uses of *xin* 心 on the one hand to mean awareness (知覺) and on the other to mean a capacity (本心), see Chen (1991, p. 31) for discussion. It may be that *liangzhi* can be used as a noun in the first meaning described in this paragraph, though I don't know of a clear example which must be interpreted in this way.

[49] There is an important and difficult question about how Wang understood the metaphysics of the referent of *liangzhi* in this capacity use. On one model of *liangzhi*, which I elsewhere call the *Activity Model*, *liangzhi* is nothing over and above what is said to issue from it: *liangzhi* would then be simply a set of events of awareness, emotions, inclinations and so on. On the *Faculty Model*, by contrast, *liangzhi* is itself a faculty or capacity, which is not identified with these events of awareness, emotions or inclinations, but is responsible for them, much as (according to some) the faculty of sight is not identical with episodes of seeing but is responsible for them. I address this difficult question elsewhere ("Examining Wang Yangming's Conscience"), but here I intend to remain neutral on it throughout.

and the expression for genuine knowledge, *zhen zhi*, 真知 – which we encountered in section 2 and which will be the target of the interpretation in the next sections.[50] *Zhen zhi* 真知 also has two characters, "genuine" *zhen* 真 and "knowledge" *zhi* 知. But in contrast to the expression *liangzhi* 良知, Wang uses the expression *zhen zhi* 真知 as a noun exclusively in the first or third senses described above: it is what one has when one genuinely knows, or the activity of genuinely knowing (i.e. of genuinely apprehending, experiencing an episode of knowledge). There is no evidence that Wang believed in a capacity of "genuine knowledge". I have paused to highlight this point because some readers of Wang seem to identify *liangzhi* and *zhen zhi*. But this is a serious mistake: in their most common uses, *liangzhi* 良知 and *zhen zhi* 真知 describe two quite different sorts of things: the one describes a capacity of *zhi*-ing, while the other describes either that which one has when one genuinely knows, or the activity of genuinely apprehending. Moreover, while I believe *liangzhi* is involved in every episode of genuine knowledge *zhen zhi*, I will argue at the end of this section that not every episode of the activity *liangzhi* is an episode of genuine knowledge. My focus in what follows will be on giving a new interpretation of *zhen zhi*, genuine knowledge. In providing this interpretation, I will not be saying anything particularly novel about *liangzhi*. Many have observed before – as I too will observe below – that an important function of *liangzhi* is knowledge of one's own mind.[51] But I believe that my claim that *zhen zhi*, genuine knowledge, is *exclusively* introspective knowledge, i.e. knowledge of one's own mind, is new.

In the following passage, Wang clearly says that *liangzhi* can know the ethical qualities of inclinations:[52]

[50] In this paragraph, when I print Chinese after an italicized Romanization, the term is to be taken as mentioned rather than used.

[51] For instance, Ivanhoe speaks of *liangzhi* as "constantly monitoring" one's mental events (Ivanhoe (1996, p. 254) and Ivanhoe (2018, p. 66)).

[52] Below, my exegesis of the unity of knowledge and action – which Wang already taught shortly after his enlightenment in Longchang in 1506 – will rely on certain doctrines about the introspective activity of *liangzhi*, which Wang reputedly began to emphasize only fifteen years later, in 1521. My interpretation might therefore seem chronologically impossible. But in the period prior to Wang's teaching about *liangzhi* he attributed many qualities similar to those he later attributed to *liangzhi* to the "original substance of the mind" (心之本體). (Especially striking parallels can be found in his 1515 *Preface to the Old Version of the* Great Learning, published in 1518, *QJ* 270-1.) He also uses *zhi* on occasion to indicate what he would later call *liangzhi*. As Wang is reported to have said himself, "From Longchang on, I have not departed from the meaning of the two characters '*liangzhi*'. But it was just that I was unable to produce these two characters in speaking to scholars, and wasted many words describing it. Now, fortunately, this meaning has been made manifest, so that in one expression, one can see clearly the whole substance." (吾『良知』二字，自龍場已後，便已不出此意，只是點此二字不出，於學者言，費卻多少辭說。今幸見出此意，一語之下，洞見全體 *QJ* 1747, reported by Qian Dehong)

[T8] 意則有是有非，能知得意之是與非者，則謂之良知。
Some inclinations are right, and others wrong; what is able to know the rightness and wrongness of inclinations is called *liangzhi*. (*QJ* 242; also translated in Ching (1972, p. 114))[53]

The words I have translated as "right" and "wrong" (*shi* 是 *fei* 非) can also mean "correct" and "incorrect". I have opted for the translations "right" and "wrong" here since Wang clearly describes the correctness as ethical or moral correctness. While there is a clear contrast between this correctness/incorrectness and goodness/badness (*shan* 善 *e* 惡), nothing of great import will hang on the exact character of this distinction in what follows. In particular, I will not be assuming that this distinction is or is not the same as the distinction between right/wrong and good/bad as that distinction is understood by moral philosophers working in English today.

In **[T8]**, Wang describes *liangzhi* as a capacity – that which is able to know the rightness and wrongness of inclinations. In other passages, however, he says not only that *liangzhi can* know the rightness or wrongness of mental episodes, but that it does:

[T9] 爾那一點良知，是爾自家的準則。爾意念著處，他是便知是，非便知非
Your *liangzhi* is your own standard. Insofar as your motivating concerns (*yinian*) are attached, it knows their rightness if they are right, and their wrongness if they are wrong. (*IPL* 206, *QJ* 105)

The expression I have translated as "motivating concerns" (*yinian* 意念) is a compound of the word I have translated "inclination" *yi* together with one which we have not yet encountered, but which I would render "concern" *nian*.[54] It is tempting to think that Wang held that something is an inclination if and only if it is a motivating concern, and that something is a motivating concern if and only if it is a concern. One

[53] It is linguistically possible to translate the above passage as "what is able to know that inclinations are right or wrong is called *liangzhi*". The difference between that translation and the one I have opted for won't be especially important in what follows (the main view could be developed using either one), but there might be some differences in detail. Obviously the translation I've chosen fits well with the emphasis on objectual knowledge argued for in n. 24. A similar point applies to the next passage, as well.

[54] This term is often rendered simply "thoughts". But this translation does not capture the fact that *nian* have an affective component; they are different from *si* (思, which I translate "thoughts") which are more often dispassionate. Liu Zongzhou (劉宗周, Jishan 蕺山, 1578–1645) for example writes (criticizing Wang) that "A thought which is set in motion by desire is a concern. Thus concerns must be eradicated although thoughts cannot be." (思而動于欲為念。故念當除而思不可除, in Wu (2007, 遺編 v. 30, 陽明心錄 3); see Chan (陳榮捷) (1983, p. 142) .) The word *nian* (which will be used as a noun in all the passages we'll discuss below) does not fit well with the concerns described by the English "to be concerned about" (as in "I am concerned about you"); it fits better with concerns described by "to be concerned with" ("He is primarily concerned with his own reputation") "to be concerned to" ("I'm concerned to get there on time") and "to be concerned that" ("I'm concerned that they aren't here yet"). Just as I noted earlier that *yi* are momentary episodes whereas inclinations are not, so too the reader should bear in mind that Wang tends to think of *nian* as momentary episodes, by contrast to concerns which can persist for a long time.

line of thought in support of this claim is that Wang uses the compound "motivating concerns" (*yinian* 意念) in some passages where one might have expected just "inclination" (*yi* 意) or "concern" (*nian* 念) (e.g. *IPL* 7 *QJ* 7, *IPL* 206 *QJ* 105), most notably in more formal essays (as opposed to conversational remarks; see **[T12]**, below). These substitutions make it tempting to think that Wang took them to be equivalent.[55] In any case, I'm going to assume for simplicity in what follows that something is an inclination if and only if it is a motivating concern. Those who do not believe that Wang endorsed this claim should see this assumption as adopted purely to simplify my exegesis: everything I will say below could be rewritten to include motivating concerns as a separate class of mental events, in addition to inclinations.

[T9] suggests that *liangzhi*'s knowledge of the rightness or wrongness of a motivating concern is automatic: it is not just that *liangzhi* is capable of knowing rightness and wrongness; it knows the rightness or wrongness of a motivating concern whenever it "is attached". In general, in Wang's idiom "being attached" would have had a negative connotation, but in this passage, Wang cannot mean that every motivating concern which is attached is thereby wrong or incorrect, since he explicitly says that they can be right or correct. It is natural instead to take Wang's discussion of motivating concerns' being "attached" simply to mean something like their being "aroused". His point is that whenever one has a motivating concern, the ethical quality of the motivating concern is known.

Wang often emphasizes the effortlessness of the exercise of *liangzhi*, providing further evidence that he believed that *liangzhi* not only could know the ethical qualities of mental episodes, but that it invariably does:

[T10] 是非之心，不待慮而知，不待學而能，是故謂之良知。
The mind which judges right and wrong (*shi fei*) does not await reflection before it knows, and does not await learning before it is able to, this is why it is called *liangzhi*. (*QJ* 1070; see Chan (1963, p. 278))

These passages together make it clear that *liangzhi* has automatic knowledge of the ethical qualities of mental events, that is:

[55] Tang (1970, 103-4) implicitly takes this position, translating *yinian* as "volitional ideas". Whatever one's views about the relationship between this compound and *yi* and *nian* on their own, it is important, I think, that it is a compound and not (as Chan (1963) sometimes renders the term) two separate nouns "thought or wish".)

There is also further evidence that inclinations and concerns have a close relationship: Wang says at least twice in the corpus that the mind is never without concerns (*IPL* 120 *QJ* 40, *IPL* 203 *QJ* 103-4). He also often describes inclinations generically as the activity of the mind (see **[T16]** below and *IPL* 174, *QJ* 86-7). Together these passages suggest that either the two are equivalent or the former is a species of the latter.

Knowing Right and Wrong If a person has a right/wrong inclination/motivating concern the person's *liangzhi* knows its rightness/wrongness.[56]

The converse of this claim should also be uncontroversial: a person's *liangzhi* can know the rightness of an inclination only if they have that inclination and it is right.

We have only seen cases where the ethical quality known by *liangzhi* is the rightness or wrongness of a mental episode. The next passage may suggest a richer range of objects, including filiality, respect and compassion:

[T11] 知是心之本體，心自然會知：見父自然知孝，見兄自然知弟，見孺子入井自然知惻隱，此便是良知不假外求。若良知之發，更無私意障礙，即所謂『充其惻隱之心，而仁不可勝用矣』。然在常人不能無私意障礙，所以須用致知格物之功勝私復理。即心之良知更無障礙，得以充塞流行，便是致其知。知致則意誠。

Knowledge is the original substance of the mind.[57] The mind is automatically able to know. When it sees one's parents, it automatically knows filiality. When it sees one's elder brother, it automatically knows respect. When it sees a child fall into a well, it automatically knows compassion. This is *liangzhi*, and should not be sought outside. If *liangzhi* is aroused, and there is furthermore no obstruction of selfish inclinations, it will be like the saying "If one fulfills one's mind which is compassionate, then one's humaneness will function inexorably." But ordinary people are unable not to have the ob-

[56] In the main text I've only considered inclinations and motivating concerns, but Wang makes related points about *liangzhi*'s knowledge of other mental events in a range of other passages. In *IPL* 290 (*QJ* 126) *liangzhi* is said to be able to be aware (*hui jue* 會覺) of emotions, even when the mind as a whole is corrupt, just as the light of the sun allows shapes and color to be distinguished when the sky is full of mist and fog. This awareness in turn is supposed to be a part of eliminating (蔽去) the bad ones; the suggestion seems to be that *liangzhi* recognizes their badness and facilitates a response to them. Wang makes more directly related points about thoughts (*si* 思) in *IPL* 169 (*QJ* 81-2); there *liangzhi* is said to be able to automatically distinguish thoughts. "Whether right (*shi* 是) or wrong (*fei* 非), warped (*xie* 邪) or correct (*zheng* 正), there is none that *liangzhi* does not know automatically."

Related points are made in *IPL* 71, *QJ* 25. Wang does not there describe this knowledge as due to *liangzhi*, but the passage likely dates from the period prior to Wang's emphasis on *liangzhi*, and it's likely he would have later held it was due to *liangzhi*.

[57] On the meaning of the expression I follow tradition in translating "original substance", see n. 7. The most natural interpretation of "knowledge" here is as referring to a capacity, discussed in n. 49. One piece of evidence in favor of this interpretation is that Wang often speaks of *liangzhi* as the original substance of the mind, and he plausibly often thinks of *liangzhi* as associated with a capacity. In many places, he clearly uses *zhi* on its own to stand for *liangzhi*; he is plausibly doing the same here. The fact that *zhi* on its own can mean *liangzhi* is in my view the key to a range of passages which are very close to a passage we will see later, **[T16]**, and which seem to conflict with KA. In these passages Wang says that knowledge comes first, and inclinations or even actions come after (*IPL* 137, *QJ* 53; cf. *IPL* 78, *QJ* 27). But I believe that in those passages, Wang is talking about the capacity for *zhi*-ing (not the activity of the capacity), and the possession of this capacity is clearly temporally (and causally, perhaps even ontologically) prior to the arising of an inclination (except perhaps for the first inclination of one's conscious life, if there is such an inclination and if *liangzhi* itself comes into being simultaneously with that inclination). So these passages are no threat to KA.

> structions of selfish inclinations. That is why they must use the practice of the extension of knowledge (*zhi zhi* 致知) and the investigation of things (*ge wu* 格物) in order to conquer selfishness and restore *li*. Then the mind's *liangzhi* will furthermore have no obstructions and will be able to operate smoothly everywhere. This then is the extension of knowledge. And if one's knowledge is extended, one's inclinations will be sincere. (*IPL* 8, *QJ* 7)

On a first reading of this passage, one might think Wang identifies knowledge of filiality with seeing the parents. But since Wang remarks explicitly that filiality (or knowledge of it) "should not be sought outside", this cannot be what he intends. It seems we should instead think of Wang's idea as follows: first, one sees one's parents. Then *liangzhi* naturally "emits" (*fa* 發) an appropriate response; presumably this response is an appropriate inclination, emotion, concern, or thought.[58] If the mental event the person produces is right (是), *liangzhi* knows its rightness, and perhaps its filiality as well; as a result, the person knows filiality.

[T11] suggests that Wang may have held that *liangzhi* knows richer qualities of mental episodes than just rightness and wrongness; he suggests that it knows filiality as well. This position would be hard to understand unless Wang thought that mental episodes like inclinations do exhibit filiality and other such qualities. But as I mentioned at the end of the previous section, Wang does not clearly say anywhere that they do. So while the passages discussed in this section make it clear that Wang held that *liangzhi* knows rightness and wrongness, and while **[T11]** makes it clear that *liangzhi* is importantly involved in a person's knowing filiality, it is not clear to me that Wang was happy to speak of *liangzhi* as knowing filiality directly. I will consider the claim that *liangzhi* knows filiality, respect, compassion and loyalty again below. But my interpretation will not require this claim.

[58] That Wang attributes the production of such mental events to *liangzhi* is clear: in *IPL* 290 (*QJ* 126), Wang says that some emotions are the functioning (*yong* 用) of *liangzhi*; in *IPL* 169 (*QJ* 81-2) he describes some thoughts as the thoughts of *liangzhi*. The fact that *liangzhi* produces emotions and thoughts, and is also responsible for recognizing whether one's own actions (i.e. mental events) are right or wrong makes it importantly similar to the conscience, which at least seems (to the puzzlement of some philosophers, see e.g. Fuss (1964, p. 118-9)) to exhibit both of these features. The suggestion that *liangzhi* be understood as "conscience" has often been made. (Early examples are: Graham (1958, p. xx), Chang (1955), Mou (1973, p. 104 n. 3). Tang (1973) often uses "conscientious consciousness"; cf. Cheng (1974). See now Bol (2008, p. 169).) But as far as I'm aware the authors who have suggested this similarity have not emphasized this crucial dimension of similarity: that, like *liangzhi*, the conscience both produces emotions and knows their rightness/wrongness; it has a "non-cognitive", affective aspect, as well as a "cognitive", epistemic one.

One interesting possible disanalogy between *liangzhi* and the conscience concerns the range of objects known by the conscience. The conscience appears to know only right and wrong, and not richer properties such as filiality. If *liangzhi* does know filiality (see next paragraph in the main text), then that would be a difference between it and the conscience. In fact it is not even clear that the conscience can know goodness/badness, but as we will see in the next section Wang clearly says that *liangzhi* does.

The passages we've examined in this section are consistent with the verbal use of *zhi* 知 meaning either "know" or "apprehend" (in the technical sense I introduced in section 2, where it means "experience an episode of knowledge"). In line with my focus on KA, as opposed to Dispositional KA, I'm going to interpret these verbal uses in the latter episodic sense, and develop my interpretation accordingly. But those who take *zhi* here to mean "know", and who (accordingly) favor Dispositional KA, should be able to take on board everything I say below, by interpreting the relevant disposition as a disposition to experience relevant episodes of knowledge in appropriate circumstances.

Wang says in many places that it is not just the *liangzhi* of sagely people which knows; a person's *liangzhi* knows the qualities of their mental events even when the person is generally morally corrupt, indeed even when they are in the midst of performing vicious bodily actions.[59] *liangzhi* automatically knows the ethical qualities of a person's ethical events, regardless of the overall state of their mind. But given that Wang held this view, he must also have held that the knowledge which *liangzhi* automatically has is not always genuine knowledge. For Wang clearly held that a person who is unfilially assaulting their parents does not genuinely know filiality. Wang's remarks to Xu Ai in **[T1]** indicate that someone who genuinely knows filiality cannot be acting unfilially. Indeed, the passage **[T11]** itself further supports the idea that *liangzhi*'s activity does not always constitute genuine knowledge. Wang does not use the expression "genuine knowledge" in this passage, but he twice qualifies his claim that one is naturally able to know filiality in a way that suggests a distinction in kinds of knowledge. He says that although people are naturally able to know filiality, in order to (genuinely) know it they must be free of selfish inclinations, and their knowledge must be extended.[60]

[59] In "The Preface to the Old Version of the Great Learning" Wang says that the original substance of the mind never fails to know (未嘗不知也 *QJ* 271). (In 1515 when he wrote this work, he did not yet speak of *liangzhi* in the way he later would.) The point is also made in his "Letter to My Younger Brothers" (*QJ* 193, translation in Ching (1972, p. 49)), and emphatically in *IPL* 152 (*QJ* 69) and *IPL* 207 (*QJ* 105). In *IPL* 206 (*QJ* 105), Wang says also that *liangzhi* knows bad inclinations, not just the good ones. (In *IPL* 169 (*QJ* 81-2), he leads by saying only that *liangzhi* can distinguish between inclinations, but here the modal is naturally interpreted to indicate that, when they arise, they are distinguished; Wang finishes the passage by saying that *liangzhi* in fact knows them all.) *IPL* 290 (*QJ* 126) is at first sight in tension with these remarks. There, Wang says that *liangzhi can* know the qualities of these states, not that it does. Moreover, in *IPL* 290 he says that an immediate consequence of *liangzhi*'s awareness of bad emotions is to eliminate them. This passage seems incompatible with the remarks just cited, since if (as Wang says in *IPL* 207 (*QJ* 105)) a thief's *liangzhi* knows the evil of what they do, and if this knowledge comes with elimination of the bad emotions, no one would be a thief. A conservative way around this problem is to see *IPL* 290 as describing genuine knowledge which, as I will describe below, does arise only if bad mental events are eliminated.

[60] In **[T11]** one could either construe the expression *liangzhi* as referring to a capacity or as referring to the exercise of that capacity. But even if one adopts the latter construal, it is clear that *liangzhi* differs from genuine knowledge: while one invariably experiences *liangzhi*-ing (in the sense of the exercise of the capacity), that *liangzhi*-ing does not invariably amount to genuine knowledge. All genuine knowledge is presumably the product of *liangzhi* (and coincident with or even identical to the exercising of that capacity),

So Wang clearly holds that the automatic knowledge of *liangzhi* is not always genuine knowledge. Our next question will be what further conditions must be satisfied if this automatic knowledge is to be genuine knowledge.

6 The obscuration argument

Wang discusses such further conditions in the following passage, where he presents what I will call the *obscuration argument*:

[T12] 故欲正其心者，必就其意念之所發而正之，凡其發一念而善也，好之真如好好色，發一念而惡也，惡之真如惡惡臭，則意無不誠，而心可正矣。。。。。凡意念之發，吾心之良知無有不自知者。其善歟，惟吾心之良知自知之；其不善歟，亦惟吾心之良知自知之。。。。。意念之發，吾心之良知既知其為善矣，使其不能誠有以好之，而復背而去之，則是以善為惡，而自昧其知善之良知矣。意念之所發，吾之良知既知其為不善矣，使其不能誠有以惡之，而覆蹈而為之，則是以惡為善，而自昧其知惡之良知矣。若是，則雖曰知之，猶不知也，意其可得而誠乎！今於良知之善惡者，無不誠好而誠惡之，則不自欺其良知而意可誠也已。

Therefore if you want to rectify your mind, you must rectify it in regard to the arousal of your motivating concerns. If, whenever a concern arises and it is good, you genuinely love it as you love lovely colors, and whenever a concern arises and it is hateful, you genuinely hate it as you hate hateful odors, then none of your inclinations will be insincere and your mind can be rectified...

Whenever a motivating concern arises, your mind's *liangzhi* automatically knows it. [If it is good] your mind's *liangzhi* automatically knows its goodness; [if it is evil], your mind's *liangzhi* also automatically knows its evil [hatefulness] …When a [good] motivating concern arises, the *liangzhi* of your mind already knows it to be good. Suppose you do not sincerely love it but instead turn away from it and eliminate it. You would then be taking good to be evil [hateful] and obscuring your *liangzhi* which knows goodness. When an [evil] motivating concern arises, the *liangzhi* of your mind already knows it to be evil [hateful]. Suppose you do not sincerely hate it but instead backslide and promote it. You would then be taking evil [hatefulness] to be good and obscuring your *liangzhi* which knows evilness [hatefulness]. In such cases one says that you know it, but in fact you do not know. How then can inclinations be made sincere? If what *liangzhi* [regards as] good or evil [hateful] is sincerely loved or hated, one's *liangzhi* is not deceived and inclinations can be made sincere. (*QJ* 1070-1, cf. Chan (1963, p. 277-9))

Wang argues that if one does not sincerely love (好) a good (善) motivating concern, then one does not know its goodness, and similarly that if one does not sincerely hate

but not every episode of the activity of *liangzhi* is an episode of genuine knowledge.

(惡) and remove (去) an evil (惡) motivating concern, one does not know its evilness. His basic argument seems to run as follows (in the case of a good motivating concern).

First, if one does not sincerely love a good motivating concern, then one turns away from it and removes it. Any loving which is not sincere involves some turning away. (Recall that "sincere" is similar in meaning to "wholehearted".) To turn away from a good inclination is to "take good to be evil" (以善為惡).

Second, Wang says that taking good to be evil obscures the *liangzhi* which knows goodness (or: "which knows that [the motivating concern] is good"). He concludes by saying that, in such a case one "says that you know it, but in fact you do not know." The passage thus advances the following argument:

1. If one turns away from and removes a motivating concern, one takes it to be bad.
2. If one takes something to be bad, one does not know its goodness.
3. So, one knows the goodness of a motivating concern only if one does not turn away from it and eliminate it.[61]

As stated explicitly in the passage, Wang's argument concerns knowledge, and not genuine knowledge. The conclusion of the argument denies that the person has knowledge *at all*. But as we have seen Wang says repeatedly in other places that – just as the sun illuminates the sky even on a foggy day – *liangzhi always* knows, even when a person is morally corrupt. These passages show that Wang can here mean only to deny that the person has genuine knowledge, since of course their *liangzhi* (even when it is obscured) has some form of knowledge. The suggestion is further supported by the end of the passage. There Wang says that in a case of motivational conflict we do *say* that the person knows, even though they do not really know. Perhaps Wang means that we say that the person knows that what they are doing is wrong, just as in Xu Ai's case, we say that the people know that they ought to be filial toward their parents. In both cases

[61] Wang's remarks naturally suggest the following argument that cognitivists about desire (i.e. those who hold that if one desires that p one believes that it's good that p) cannot allow that people want what they know to be not good:

(a) If one wants that p, one believes that it's good that p.

(b) If one believes that p, one does not know that $\neg p$.

(c) If one wants that p one does not know that it's not good that p.

Replies to this argument will naturally focus on the second premise. Cognitivists about desire sometimes distinguish between "guises" under which beliefs are held; they might claim that someone may believe that p and simultaneously know that $\neg p$ so long as the belief and the knowledge are under different guises. Fans of "fragmentation" allow that an individual may believe p and believe $\neg p$ (in different fragments) even in cases where it would be natural to describe the individual as knowing p. Here is not the place to delve any deeper into these issues.

Wang seems to recognize a divergence between his own way of speaking of (genuine) knowledge, and the form of knowledge ordinary usage tracks. So the passage should be understood to argue that genuine knowledge requires a kind of overall cognitive harmony, while conceding that other forms of knowledge may not.

The second premise of the obscuration argument gives us insight into why Wang thought genuine knowledge was a higher form *of knowledge* than ordinary knowledge. The passage suggests that *liangzhi*'s knowledge of the goodness of a motivating concern is a lesser form of knowledge if it is "obscured" by something like a conflicting belief ("taking good to be bad").[62] We can illustrate the idea further with an example. Consider a person who, whenever they want to walk from their home to a particular temple, can get there without any trouble, but who, whenever they want to instruct others how to get from their home to the temple, invariably sincerely denies that the temple can be reached in the way they themselves walk there. If there could be any people who both know a claim, and believes the negation of that claim, this character is a good candidate: they know – as evinced by their ability to walk to the temple – that the temple can be reached from their home by walking along a particular street, but they also believe – as evinced by their speech – that the temple cannot be reached by this route. Now compare this person ("the conflicted person") to someone ("the unconflicted person") who can both walk from their home to the temple and instruct others correctly about how to reach it. As applied to this case, Wang's thought seems to be that the conflicted person has a lesser form of knowledge than the unconflicted person, precisely because of the conflict between the conflicted person's beliefs and what they know. Wang's basic thought does not depend on the claim that there are better or worse forms of knowledge. We could instead paraphrase it by saying that the unconflicted person's state of mind is better in a distinctively epistemic or doxastic respect than the conflicted person's state of mind, precisely because the unconflicted person does not suffer from doxastic conflict. In the obscuration argument Wang seems to indicate the "betterness" of the one form of knowledge by describing the unconflicted person's knowledge as "not obscured". The discussion of the unity of knowledge and action in **[T1]** suggests that he described this form of knowledge using the honorific "genuine".

The first premise of the obscuration argument connects loving and "taking to be good". It is natural to think of this premise as connecting a certain affective response ("loving") to something like a doxastic feature of a person's mental state ("taking").

[62] This belief could of course be implicit; it is not assumed to be conscious, and its formation is not assumed to be effortful.

This connection then underwrites the claim that genuine knowledge – which is elevated above ordinary knowledge in the first instance by the distinctively epistemic or doxastic characteristic described in the previous paragraph – also enjoys a more intimate connection to action. If a person is to genuinely know the goodness of a good motivating concern, then they must not take it to be bad. But if they do not love it, then they in some sense turn away from it, and so take it to be bad. So, being free from doxastic conflict requires being free from motivational conflict as well. And (as we saw in the lead-up to Inclination Action), when a person is free from motivational conflict, and has an inclination/motivating concern to perform a good action, then they are performing that good action. So, if a person genuinely knows, they are free from doxastic conflict; if they are free from doxastic conflict, they are free motivational conflict; and if they are free from motivational conflict and and have an inclination to perform a good action, then they are performing that good action.

Just before the passage in the *Great Learning* which discusses loving a lovely color (which Wang quotes in **[T1]**), the text says that having sincere inclinations means not deceiving oneself ("what is called making one's inclinations sincere, is to have no self-deception" 所謂誠其意者，毋自欺也). In **[T12]** and other passages, Wang invokes this traditional connection between sincerity and lack of self-deception to cement the connection between lack of motivational conflict and lack of doxastic conflict.[63] At the close of the passage, Wang says that a person's inclinations are sincere (a condition of motivational wholeheartedness) if (and presumably only if) the person is not deceiving themselves (自欺) (a condition of doxastic coherence). The passage makes clear that Wang takes a person to deceive themselves in the relevant way if they have a response to an inclination which is not appropriate to its ethical quality (i.e. loving a bad inclination; hating a good inclination), and that he holds that this inappropriate response amounts to "taking" the inclination to have an ethical quality it does not have. Wang seems to see his argument as elucidating the connection between sincere inclinations and freedom from self-deception that he found in the *Great Learning*.

In **[T12]**, Wang talks about *liangzhi*'s knowledge of goodness/badness (*shan* 善, *e* 惡), but does not mention*liangzhi*'s knowing rightness/wrongness or correctness/incorrectness (*shi* 是, *fei* 非). Our next passage draws a connection between *liangzhi*'s loving/hating and its knowing rightness/wrongness which suggests that Wang would accepted the same paradigm for knowledge of rightness, as I have suggested he endorsed for knowledge of goodness:

[63] We saw this same point also in **[T7]**; cf. *IPL* 171, *QJ* 84.

[T13] 良知只是個是非之心，是非只是個好惡，只好惡就盡了是非，只是非就盡了萬事萬變。

Liangzhi is just the mind which [judges] right and wrong (*shi fei*). [Judging] right and wrong (*shi fei*) is just loving and hating. If you have just loved and hated, then you have exhausted [judging] right and wrong (*shi fei*). If you have just [judged] right and wrong (*shi fei*), then you have exhausted the ten thousand affairs and changing [circumstances]. (*IPL* 288, *QJ* 126)

The characters for the adjectives "right" and "wrong" (*shi fei* 是非) (which are the same as the nouns "rightness" and "wrongness") can each also be used as verbs, meaning "judge to be right" or "approve" on the one hand and "judge to be wrong" or "disapprove" on the other. In this passage Wang says that for *liangzhi* to perform these activities is for it to love or hate. This strongly suggests that Wang would have extended the paradigm described in **[T12]** to *liangzhi*'s knowledge of rightness/wrongness: just as we have seen that loving/hating are connected to *liangzhi*'s knowledge of goodness and badness, Wang espouses a similar connection between loving/hating and recognition of rightness/wrongness. [64]

The obscuration argument shows that Wang took genuine knowledge to differ from ordinary knowledge in what we can recognize as a distinctively epistemic or doxastic

[64]The passage also raises a new question about the relationship between *knowledge* of goodness/rightness and loving. What I'll call a *cognitivist* theory of the knowledge of *liangzhi* holds that loving and hating are separate activities that go along with knowing the quality of a mental event (e.g. a motivating concern) but are not identical to that knowledge. A *noncognitivist* theory of the knowledge of *liangzhi*, by contrast, holds that loving/hating and knowing rightness/wrongness are identical; for *liangzhi* to know the rightness of a right motivating concern just is for it to love that motivating concern (and similarly for wrongness and hate). Angle & Tiwald (2017) endorse something similar to a non-cognitivist interpretation; according to them, *liangzhi* (taken here to indicate the exercising of the capacity) is a "category" "as much a kind of emotion as it is a cognitive judging state" (p. 104). This position can be further supported by independent evidence that Wang believed that all episodes of knowledge are inclinations (*IPL* 201, *QJ* 103; *IPL* 174 (*QJ* 86-7), alongside evidence that Wang held that loving and hating are inclinations (*IPL* 101, *QJ* 34). (Shun (2011, p. 103-4) also calls loving and hating here inclinations.) I don't know of texts which decide between these two positions, though I slightly favor the latter.

There might also seem to be at least two different ways to think of the sincere loving Wang discusses in **[T12]**. On a *harmony interpretation*, to love a motivating concern is simply for there to be no other mental events which conflict with it; to hate it is for there to be some conflicting event. This characterization of loving and hating would make them exhaustive alternatives; turning away from would then be identified with hating, although promoting might not be identified with loving. On a different, *metacognitive interpretation*, the loving (hating) of a good (bad) motivating concern is a mental event (perhaps an inclination in its own right) on a par with the "first-order" motivating concern; whether the loving or hating is sincere depends on whether it is harmonious with the rest of what is going on in one's mind. This allows that turning away from could be distinct from hating.

But the discussion in **[T13]** favors the metacognitive interpretation fairly strongly, and indeed seems to rule out the harmony interpretation. For when Wang says that *liangzhi* judges right and wrong, he is clearly imagining that there can be loving without sincere loving. This loving therefore cannot be identified with the lack of mental conflict (otherwise it would be sincere!), so it suggests that Wang thinks that loving can occur as a mental episode in its own right.

way. He argues moreover that a person can satisfy the extra (doxastic) conditions for genuine knowledge only if they are free from motivational conflict. While he presents this argument as applied to knowledge of goodness/badness in **[T12]**, **[T13]** suggests he would have endorsed a similar paradigm for knowledge of rightness/wrongness as well. Wang's discussion in these passages paves the way for a connection between genuine knowledge and action, by way of the connection between freedom from doxastic conflict and freedom from motivational conflict. In the next section I develop this connection further.

7 The introspective model

The obscuration argument suggests that Wang held that for relevant F:

Knowledge Inclination A person genuinely knows Fness if and only if their *liangzhi* knows the rightness or goodness of an inclination to perform an F action, and that inclination is sincerely loved.

This claim – together with Inclination Action and Knowing Right and Wrong – suffices to explain Wang's belief in Unity.

To see this, let us assume (as is strongly suggested by **[T12]**) that a good inclination is sincere if and only if it is sincerely loved. By Inclination Action, a person is acting filially if and only if they have an inclination to perform a filial action and that inclination is sincerely loved. By Knowing Right and Wrong (and its obvious converse), a person has an inclination to perform a filial action – which we may suppose is a right or good inclination – if and only if their *liangzhi* knows the rightness/goodness of this inclination. So, a person has an inclination to perform a filial action and that inclination is sincerely loved if and only if their *liangzhi* knows the rightness of the inclination and the inclination is sincerely loved. By Knowledge Inclination, a person's *liangzhi* knows the rightness of an inclination to perform a filial action and it is sincerely loved if and only if they genuinely know filiality. So, a person is acting filially if and only if they genuinely know filiality.

Knowledge Inclination, Knowing Right and Wrong and Inclination Action entail Unity against very natural background assumptions. But Knowledge Inclination is still too weak to constitute an interesting thesis about genuine knowledge. Knowledge Inclination describes the conditions under which a person has genuine knowledge, but it does not tell us what genuine knowledge is. It leaves open the possibility Wang held that the event of genuinely knowing filiality just is the event of acting filially. But this

is precisely the kind of claim that would make Wang's thesis analogous to the misleading advertising about the elixir of eternal life I parodied in the introduction. If Wang defined genuine knowledge as the event of acting filially, he would simply have stipulated the truth of his doctrine, and would have given no explanation of why these actions should count as a form of knowledge. In the remainder of this section, I will suggest that Wang endorsed a further claim about the nature of genuine knowledge. This further claim will entail Knowledge Inclination, but it will rule out identifying genuine knowledge with action; it will not merely say under what conditions a person has genuine knowledge, but also say which mental events are episodes of genuine knowledge.

To clarify the difference between the question of what genuine knowledge is and the question of when one has genuine knowledge, it may be useful to consider first an incorrect interpretation, which I will call the *simple* model. There is some evidence that Wang believed that all episodes of knowledge are inclinations (*IPL* 201, *QJ* 103; *IPL* 174 (*QJ* 86-7). For the purposes of developing this simple model, let's assume that they are. The key claim of the simple model is:

Simple Knowledge Something is an episode of a person's genuinely knowing Fness if and only if it is an inclination of theirs to perform an F action and that inclination is sincerely loved.

On the plausible assumption that a person genuinely knows Fness at a time if and only if something is an episode of that person's genuinely knowing Fness at that time, Simple Knowledge entails Knowledge Inclination: the theses agree on the conditions in which genuine knowledge occurs. But Simple Knowledge does not just describe the conditions under which genuine knowledge arises. It identifies genuine knowledge with a particular mental event, in this case an inclination to perform an F action. So, for instance, in the case of someone who genuinely knows filiality, the proponent of this thesis would claim that the person's inclination to perform a filial action is an episode of genuinely knowing filiality. Of course not all inclinations to perform filial actions are episodes of genuinely knowing filiality – one might have a filial inclination and turn away from it – but in this case the filial inclination would be sincerely loved, so it would be an episode of genuinely knowing filiality.[65]

[65] The connection between knowledge and inclinations, together with the principle Mental Action (see n. 45), suggests a form of "unity of knowledge and action" that was not in view earlier in the paper, namely that for relevant F:

Instance KA Every episode of knowing Fness is an action.

If Wang endorsed Mental Action, and held that every instance of knowledge is an inclination, then he held

There is strong evidence that Wang did not endorse the simple model. Wang is clear that *liangzhi*'s knowing is distinct from the event of having a (first-order) motivating concern, thought or inclination. In **[T9]**, for example, Wang clearly speaks of the knowledge of *liangzhi* as an event distinct from having a right or wrong motivating concern. In *IPL* 169 (*QJ* 81-2), after discussing thoughts that are the "aroused functioning" of *liangzhi* (良知之發用), Wang says that "*liangzhi* also is automatically able to know" (良知亦自能知得). This "also" strongly suggests that Wang takes the knowledge to be distinct from the thoughts *liangzhi* produces. Wang has ample opportunity to say that what it is to know the rightness of a right motivating concern just is to have that motivating concern. But he doesn't.[66] On the natural assumption that Wang is talking about genuine knowledge in **[T12]**, he connects the knowledge of *liangzhi* directly to genuine knowledge in that passage. There, he repeatedly speaks of a motivating concern arising on the one hand, and *liangzhi* automatically knowing that motivating concern on the other. His language again strongly suggests that these are distinct events. He says that (e.g.) turning away and eliminating a good concern would obscure *liangzhi* which knows goodness. His point is that this distinct mental event of knowing goodness could not in the relevant circumstances lead to or constitute genuine knowledge. On the extremely natural supposition that *liangzhi*'s knowing the goodness of a filial concern is

that every instance of knowledge (whether genuine or not) is an action. On the supposition that genuine knowledge is a form of knowledge, it would follow that every episode of genuinely knowing Fness is also an action. I don't know of passages in which Wang himself states this claim, but given Mental Action and the claim that episodes of knowledge are inclinations, he would be committed to it on any reasonable interpretation of what he says, including my own. Perhaps this claim could help to explain what Wang had in mind in some of the passages I discuss in appendix A, where he seems to describe a very intimate connection between knowledge and action.

Some deny that in *IPL* 201 (*QJ* 103) (and related passages) Wang claims that all instances of knowledge are inclinations. For instance Chen (1991, p. 168) emphasizes the distinction between *liangzhi* and inclinations that Wang seems to draw in his letter to Wei Shishuo: "Inclinations and *liangzhi* must be distinguished very clearly. Whenever one responds to an object and produces a concern, these are all called inclinations. Some inclinations are right, and others wrong; what is able to know the rightness and wrongness of an inclination is called *liangzhi* (意與良知當分別明白。凡應物起念處，皆謂之意。意則有是有非，能知得意之是與非者，則謂之良知。*QJ* 242; also translated in Ching (1972, p. 114)). But this passage does not force the distinction Chen sees in it; it is most naturally read as implicitly restricted to *selfish yi*, and thus as distinguishing the activities of *liangzhi* only from selfish *yi*. Wang often describes selfish desires without adding the qualification "selfish" explicitly (e.g. *IPL* 290, *QJ* 126). So all he is saying here is that not all inclinations are episodes of knowledge (not, as Chen thinks, that not all episodes of knowledge are inclinations). The discussion of Angle & Tiwald (2017, p. 105-6) might be read to suggest (similarly to Chen) that *yi* cannot be the functioning of *liangzhi* ("intentions [*yi*] and good knowing [*liangzhi*] are structurally different from one another"). But in personal communication, they agree with the point just made that all that follows is that the activity of *liangzhi* can't be *selfish yi*.

[66]The one text we've considered that is *consistent* with the claim that they're identical is **[T11]**. But consistency isn't strong evidence: Wang just isn't concerned there with identifying exactly which mental event is knowledge. He's focused on the fact that *liangzhi* is responsible for the appropriate response to one's situation (and how that is relevant to one's knowledge of filiality), not on its role in recognizing the quality of that response.

a component of genuine knowledge of filiality, these passages show that the simple model is incorrect.[67]

These arguments against the simple model suggest that Wang identified genuine knowledge of filiality not with an inclination to perform a filial action, but with *liangzhi*'s introspective knowledge of the rightness or goodness of such an event. This position is in any case extremely natural. Wang emphasizes *liangzhi*'s role in knowing the ethical qualities of mental events. He also connects *liangzhi* to elevated cognitive achievement (e.g. in **[T11]**) and in relation to the unity of knowledge and action (e.g. in **[T6]**). It is natural to suspect he would have connected these dimensions of *liangzhi*, and that he would have endorsed:

Introspective Knowledge Something is an episode of a person's genuinely knowing Fness if and only if it is an episode of their *liangzhi*'s knowing the rightness or goodness of an inclination to perform an F action which is sincerely loved.[68]

Once again, under the natural assumption that a person genuinely knows Fness at a time if and only if something is an episode of that person's genuinely knowing Fness at that time, Introspective Knowledge entails Knowledge Inclination. But Introspective Knowledge does not just describe the conditions under which genuine knowledge arises. It identifies genuine knowledge with a particular mental event, in this case *liangzhi*'s knowledge of the rightness of the person's relevant inclination.

I earlier mentioned the possibility that *liangzhi* knows richer objects than just rightness and wrongeness, such as filiality, respect and compassion. If one believes that Wang did accept that *liangzhi* could know these richer objects, then one could accept more directly that the knowledge that *liangzhi* has of the filiality of an inclination to

[67] The simple model is attractive as an interpretation of some of Wang's remarks. For instance, it seems to me a possible understanding of the important passage of the letter to Gu Dongqiao (*IPL* 132, *QJ* 46-7) taken in isolation. In this paper I'm interested in exploring the extent to which a coherent picture can be attributed to Wang Yangming. But if I were forced to attribute one inconsistency to him which would lead to the most powerful explanation of his remarks, I would say that he was sometimes drawn to the simple model, and sometimes to the introspective model. This would explain in a single fell swoop the apparent tension between passages in which he subscribes to Identity and those in which he says things that are incompatible with it.

[68] There is a question whether Wang thinks people can have genuine knowledge of bad ethical qualities. It is clear that Wang thinks *liangzhi* knows wrongness just as much as it knows rightness, and in fact the introspective model as I'll present it is compatible with Wang thinking there can be genuine knowledge of badness (I'll only develop it for the example of filiality). But allowing genuine knowledge of badness would allow that both KA and AK could fail, if the F were replaced with "wrongness": the conditions under which a person would have genuine knowledge of badness are naturally taken to be conditions under which the person was in fact acting rightly, by sincerely hating a bad motivating concern. Since this conflicts with the general picture Wang seems to be developing, I think it's better to assume that genuine knowledge of wrongness or badness isn't in view.

perform a filial action is genuine knowledge of filiality if and only if that inclination is sincerely loved. This alternative interpretation has the attraction that it postulates a more intimate connection between the objects of *liangzhi*'s knowledge, and the objects of genuine knowledge. But as I have said, the claims that Wang held that inclinations instantiate qualities such as filiality and that *liangzhi* knows the filiality of inclinations are not particularly well supported by the texts.

A simple worked example may help to further illustrate the differences between Simple Knowledge and Introspective Knowledge. Suppose a person sees their parents, and as a result has an inclination to perform a filial action, for example, the action of cooling their parents. Everyone should agree that this person's *liangzhi* knows the rightness or goodness of the inclination to cool their parents. (Recall that I said I wouldn't be presenting particularly controversial views about *liangzhi*.) But let us suppose also that the person also sincerely loves this inclination, so that they are cooling their parents, and this bodily action is filial (so, if Wang endorsed AK, they genuinely know filiality).

Our two theses concern which mental event of this person is their genuine knowledge of filiality. According to Simple Knowledge (which we saw to conflict with the texts) it would be the inclination to perform a filial inclination itself. According to the introspective model, by contrast, the genuine knowledge of filiality is the event of *liangzhi*'s knowing the rightness of the filial inclination. On this model, Wang claims that a separate event of knowledge beyond the filial inclination is genuine knowledge.[69]

Inclination Action, Knowing Rightness/Wrongness and Introspective Knowledge make up my introspective model of the unity of knowledge and action. Since Introspective Knowledge entails Knowledge Inclination under natural assumptions, these three thesis vindicate Unity in the same way as was outlined at the beginning of this section. But Introspective Knowledge has two important consequences which Knowl-

[69]There is a third thesis, Total Knowledge, which I consider at length in the companion paper, "Perception and Genuine Knowledge in Wang Yangming". According to Total Knowledge, the genuine knowledge is the total mental event of the person at the relevant time, i.e. the event composed of all ongoing mental events. This total event includes the filial inclination, along with the knowledge of *liangzhi*, along with any other mental events ongoing at that time.

Total Knowledge contrasts with both Simple Knowledge and Introspective Knowledge in an important respect. It allows for the claim that perception may be a component of certain instances of genuine knowledge. The introspective model precludes this claim: even if the person in our example is perceiving their parents, that perception will not be a *component* of the genuine knowledge according to the introspective model, since on this model the knowledge is identified with *liangzhi*'s knowledge of the rightness of an inclination. But many will reject this consequence of the introspective model, holding that perception *is* a component of genuine knowledge, at least sometimes. See, for remarks along these lines, Nivison (1973, 132) (reprinted in Nivison (1996b, p. 243)), Nivison (1973, p. 134) (reprinted in Nivison (1996b, p. 244)), (Ivanhoe, 2002, p. 99), Ivanhoe (2009, p. 113), Ivanhoe (2011, p. 274), Angle (2005), Angle (2010). (I am not here claiming these authors endorse this view; detailed discussion of that question is left to the companion paper.)

edge Inclination does not have.

First, against natural background assumptions, Introspective Knowledge rules out the thesis Identity, which I introduced in section 2. (Identity is consistent with Knowledge Inclination.) According to the introspective model, a person's genuine knowledge of filiality is their *liangzhi*'s knowledge, and this knowledge will not in general be identical to every filial action the person is then performing. For instance, if a person is cooling their parents, the bodily action of cooling their parents is not identical to their *liangzhi*'s knowing the rightness of their (sincere) inclination to cool their parents.[70]

The fact that Introspective Knowledge is incompatible with Identity is closely connected to the way that, according to the introspective model, Wang did not simply stipulate the truth of Unity. Given Introspective Knowledge, genuine knowledge is recognizably a form of knowledge (or: a knowledge episode): it is completely natural to class *liangzhi*'s apprehension of the ethical qualities of one's own mental events as a form of *zhi*, knowledge. Wang is thus not guilty of stipulating that some arbitrary feature of virtuous action is to be called knowledge. Rather, as we saw in the previous section, he presents an argument that certain instances of knowledge count as privileged in a distinctively doxastic or epistemic way. He then advances a substantive thesis about the connection between the way in which these episodes are doxastically privileged and the absence of motivational conflict for the person overall.

A second important consequence of Introspective Knowledge is that it rules out the claim that perception of the external world is a component of an episode of genuine knowledge.[71] A person's perceiving the world around them is a different event from their *liangzhi*'s knowledge of the ethical qualities of their inclinations. Their perceiving of the world around them is also not *part* of the event of *liangzhi*'s knowing the quality

[70] This point holds even if we accept Mental Action. For even on this view, one of their filial actions at that time, for instance, is their filial inclination, concern, emotion or thought (again, in the right circumstances), which is distinct from *liangzhi*'s knowledge of its filiality.

Note that if Wang thought that mental events were part of one's action (regardless of whether he accepted Mental Action) Wang could accept:

All Knowledge is Part of Action Every episode of genuinely knowing filiality is part of a filial action.

Again, like Instance KA (see above n. 65), this may help to explain some of the remarks discussed in Appendix A, where Wang seems to describe a more intimate connection between knowledge and action.

[71] In the next three paragraphs, I use "perceiving the external world" to indicate that I am making claims about perception through the usual five senses, seeing, smelling, hearing, tasting and touching. One might of course describe *liangzhi*'s knowledge of the ethical qualities of one's mental events as arising from a form of "inner sense", and call that by extension a form of perception. I have no qualms with such a redescription; all I have in view here is the claim that perception through the five senses is not a component of genuine knowledge. Note that I also believe Wang held that *liangzhi* was responsible for some perception of the external world. Nothing I say in what follows conflicts with this: all I am claiming is that this perception of the external world would be a different event from the (internal) recognition of the ethical qualities of a mental event.

of their inclinations. The claim that genuine knowledge does not have perception of the environment as a component of it is of course consistent with the claim that perception of the world is a part of the causal process that led to this knowledge, and even the claim that a person could not have this knowledge unless they perceived the external world. Neither of these claims entails that the perception was a component of the event of their (genuinely) knowing the relevant ethical qualities.

Since Introspective Knowledge entails that perceiving is no part of genuine knowledge, it entails that seeing a lovely color or smelling a hateful odor cannot themselves be episodes of genuine knowledge (or even parts of such episodes). This claim will be controversial: the norm among interpreters has been to assume that these are the *paradigmatic* examples of genuine knowledge.[72] I cannot undertake here a detailed defense of my position on this text; a more extended defense can be found in the companion paper. Here I will make just two points. First, the text of **[T1]** does not force on us the claim that these are examples of genuine knowledge. They are introduced by the word *ru* 如, which *can* mean "for example", but can also (equally naturally) mean "like", introducing an analogy. And there is reason to think even in that passage alone that Wang intended "like" and not "for example". As I noted in n. 16, there are important disanalogies between the way in which Wang talks about these examples, and the way he talks about clearer cases of genuine knowledge. Whereas Wang talks about knowing filiality or respect (the qualities), he doesn't talk about knowing the loveliness of the color or the hatefulness of the odor; instead he talks about knowing the color and the smell. Wang may have set up this disanalogy deliberately to highlight that the knowledge of the color or the smell are not themselves examples of genuine knowledge.[73]

[72] Cua (1982, p. 9-14) breaks this norm; he explicitly treats the examples of both the color and odor as analogues, and holds quite generally that the only examples of genuine knowledge are examples like genuine knowledge of filiality. I differ from Cua in allowing that the example of cold pain and hunger could be examples of genuine knowledge; see n. 73.

[73] The examples in *IPL* 132 (*QJ* 46-7) of knowing one's soup, knowing one's clothes, and knowing the road one will travel on, also seem to be analogues for, and not examples of genuine knowledge. There, Wang does his best to show how his theory of knowledge and action can fit with Gu Dongqiao's cases. And while he certainly believes that knowledge and action are closely connected in those cases in a way that Gu Dongqiao had not appreciated, it is not forced on us to say that Wang in fact takes them to be examples of genuine knowledge.

Does this mean that on the present view one can only have genuine knowledge of (good) ethical qualities such as filiality or respect? The question of whether there can be genuine knowledge of non-moral matters has been much discussed in the literature (e.g. Frisina (1989) and Cua (1982)). Fortunately, the interpretation here does not commit us to taking a stance on it. For although the examples from **[T1]** and *IPL* 132 (*QJ* 46-7) turn out not to be examples of genuine knowledge, it is open whether there are other non-moral examples of genuine knowledge. Plausible candidates for such examples are those of hunger, cold and pain given in **[T2]**. Those examples are Wang's own – not forced on him by a classic text, or by

Second, we are now in a position to see why Wang thought these examples were so important in illustrating key features of genuine knowledge, even though they were not themselves cases of genuine knowledge.In **[T12]** Wang goes out of his way to compare sincere love for a good motivating concern to love for a lovely color. Just as when one sees a lovely color one automatically loves it and when one smells a hateful odor one automatically hates it, so too when *liangzhi* knows the ethical quality of a mental event, it automatically loves or hates it. So even though the examples of the color and the odor aren't examples of genuine knowledge, the basic structure of the connection between perception and loving and hating they describe is perfectly mirrored in *liangzhi*'s introspective knowledge of and metacognitive response to the ethical qualities of mental events. This precise relationship makes them excellent paedagogical tools for describing genuine knowledge, even though they are not examples of it.[74]

8 Conclusion

I argued in section 2 that the core of the unity of knowledge and action can be understood as the claim:

Unity A person genuinely knows Fness if and only if they are acting Fly.

In section 3, I argued that Wang believed that, for relevant F:

Mind Action A person is acting Fly if and only if they have a mind which is F.

I also discussed in detail the possibility that Wang endorsed the view that the ethical quality of an action is entirely determined by the state of mind of the person who performs the action. In section 4, I argued for a further claim about the connection between mental events and actions:

Inclination Action A person is acting Fly if and only if they have an inclination to perform an F action and that inclination is sincere.

In section 5, I argued that Wang accepted:

an interlocutor's letter – and they have the basic components of genuine knowledge I have discussed. In these cases, there are not clear candidates for the kind of obscuring response discussed in **[T12]**. But this is not a deep issue. It may merely mean that all introspective knowledge of hunger is safe from responses that would prevent it from counting as genuine knowledge.

[74] In fact, if one endorses the non-cognitivist interpretation of the knowledge of *liangzhi* (see n. 64), the connection is even more intimate: loving a good motivating concern would be identical to genuine knowledge of its quality if and only if the loving is sincere.

Knowing Right and Wrong If a person has a right/wrong inclination/motivating concern the person's *liangzhi* knows its rightness/wrongness.

That Wang accepted the converse of this claim requires no argument: clearly a person's *liangzhi* can know the rightness of an inclination only if the person has that inclination and it is in fact right.

In sections 6-7 I proposed that Wang furthermore held:

Introspective Knowledge Something is an episode of a person's genuinely knowing Fness if and only if it is an episode of their *liangzhi*'s knowing the rightness or goodness of an inclination to perform an F action, and that inclination is sincerely loved.

Inclination Action, Knowing Right and Wrong and Introspective Knowledge together explain how Wang accepted Unity. Under the natural assumption that a person genuinely knows Fness at a time if and only something is an episode of that person's genuinely knowing Fness at that time, Introspective Knowledge entails that a person genuinely knows filiality if and only if their *liangzhi* knows the rightness of an inclination of to perform a filial action, and that inclination is sincerely loved. By Knowing Right and Wrong (together with its obvious converse), a person's *liangzhi* knows the rightness of an inclination to perform a filial inclination if and only if they have an inclination to perform a filial action. By Inclination Action, a person has an inclination to perform a filial action and that inclination is sincerely loved if and only if they are acting filially. So, a person genuinely knows filiality if and only if they are acting filially.

Introspective Knowledge also explains (as Knowledge Inclination did not) how genuine knowledge is rightly called a form of knowledge. *Liangzhi*'s recognition of the rightness of inclinations is very naturally described as knowledge (or: an episode of knowledge). Moreover, Wang does not simply stipulate that an unnatural subclass of this form of knowledge be called "genuine". The obscuration argument shows that he believed genuine knowledge was privileged in (what we might describe as) a distinctively epistemic or doxastic way. A person who genuinely knows does not suffer from doxastic conflict – they are not deceiving themselves – whereas a person who knows but does not genuinely know, does suffer from such conflict. These two claims – that genuine knowledge is a form of knowledge, and that it is elevated on distinctively epistemic grounds – are the key to my reply to the criticism of the unity of knowledge and action I described in the introduction. On the interpretation I have presented, Wang was not simply stipulating that genuine knowledge was identified with action.

One might caricature the introspective model by saying that genuine knowledge is an inefficacious and intrinsically valueless prize superadded to the victory of virtuous action. The caricature has a seed of truth in it. According to the introspective model genuine knowledge is not an ingredient in a reasoned process of deliberation; it is an automatic recognition of the virtuousness of the (mental events which lead to the) action one is performing. But the caricature is also unfair: if genuine knowledge is a prize superadded to the victory of virtuous conduct, it is a special kind of prize that cannot be taken by theft, deceit or foul play. The prize comes when and only when the victory is honestly won. So in a certain sense, as Wang emphasizes again and again, to aim at virtuous action just is to aim at genuine knowledge; the cultivation of knowledge and action are one and the same (see appendix A for further discussion). In this sense to describe the prize as valueless or inefficacious would be to describe the victory in the very same terms.[75]

The caricature does, however, help to illustrate how starkly Wang differed from his predecessors as he understood them. According to Wang, Cheng Yi and Zhu Xi had held that ethical knowledge facilitates virtuous action in part through its role in deliberation. According to this Cheng-Zhu orthodoxy as Wang understood it, knowledge came first and virtuous action later.[76] According to the caricature, Wang replaced this knowledge-first position with one that held that knowledge came last. While I've said that this isn't quite fair in detail, it is a helpful way of thinking about the difference between the two positions.

A comparison to a kind of coherentism may help to further bring out some key aspects of Wang's moral epistemology. *Liangzhi* invariably produces good inclinations, emotions and so on. *Liangzhi* also invariably knows the goodness of these events and loves them. To transform *liangzhi*'s knowledge of this goodness to genuine knowledge of it, one must simply make the rest of one's ongoing mental events cohere with these products of *liangzhi*, so that the loving can be sincere. This coherence is achieved by eradicating selfish inclinations which conflict with the underlying products of *liangzhi*, including any which count as "turning away" from it. Making one's inclinations sincere and eliminating self-deception are both a matter of achieving overall cognitive harmony. In this regard, the achievement of genuine knowledge resembles the cognitive

[75] On the non-cognitivist model of *liangzhi*'s knowledge of the rightness of inclinations (see n. 64), the knowledge of *liangzhi* is identical to loving or hating. So the knowledge is itself related to the promotion of the inclination, though not on its own sufficient to generate action.

[76] Zhu is mentioned explicitly in *IPL* 133, *QJ* 48 just before **[T6]**, so that it is clear he is the target of that critique. Cf. *IPL* 5, *QJ* 5 (今人卻就將知行分作兩件去做，以為必先知了然後能行，我如今且去講習討論做知的工夫，待知得真了方去做行的工夫，故遂終身不行，亦遂終身不知。) and *QJ* 1331 (近世學者分知行為兩事，必欲先用知之之功而後行，遂致終身不行，故不得已而為此補偏救弊之言。), *QJBB* 173.

achievement of acquiring justified belief as imagined by a simpleminded coherentist: so long as one's worldview coheres the beliefs which are a part of it are justified. But unlike a simpleminded coherentist, Wang holds that people's beliefs are anchored to reality by the deliverances of *liangzhi*. Since *liangzhi* is ineradicable, one's overall mental state can be coherent or harmonious only if it coheres with the deliverances of *liangzhi*. Since *liangzhi* is infallible, coherence with *liangzhi* guarantees virtuous conduct (and knowledge of the relevant virtue). Self-cultivation can be simply described as a matter of achieving cognitive harmony because the only possible form of harmony is harmony with what is right: if one's web of belief is well enough laced with whispers from god, it is guaranteed to cohere if and only if the beliefs are true.

A common reaction to Wang's moral epistemology is that the doctrine of the infallibility of *liangzhi* is implausible. I agree that something is implausible – or at least, highly controversial – here, but the infallibility of *liangzhi* does not strike me as the problem. The conscience is also infallible in its judgments and unerring in what it prompts one to do and feel. One's conscience can't tell one that an action is right if it is wrong or be a source of regret for a good action. It may seem at the time that one's conscience gives verdicts that later turn out to be inapt, but this only shows that something can seem to be the deliverance of one's conscience when it isn't. In fact, although sometimes Wang speaks of *liangzhi* in ways that suggest it goes along with a distinctive phenomenology, when pressed (e.g. *QJ* 242, Ching (1972, p. 114)) he shies away from this style of view: just as in the case of the conscience, one can't rely on phenomenology (or for that matter anything) to distinguish what's *liangzhi* from what isn't. There's nothing implausible about the infallibility of the conscience, anymore than there is anything implausible about the claim that if one knows something, then it is true: it is part of the characterization of something's being a deliverance of one's conscience that it is correct – that it is true if it is a judgment and apt if it is an emotion. So it is not the infallibility of *liangzhi* that seems to me implausible. What does seem implausible, however, is the conjunction of the claim that *liangzhi* is infallible with the claim that *liangzhi always* delivers a verdict or produces a response to one's mental events. Most people often have the experience that their consciences are silent. It is hard to believe that this apparent silence is merely due to the blaring white noise of our selfish desires.

A person is *akratic* if they know that they ought to perform an action, but they voluntarily fail to do it nevertheless. Wang's view of the unity of knowledge and action is sometimes said to amount to the denial of the possibility of *akrasia*. Wang does hold that a person genuinely knows filiality if and only if they are acting filially. So he denies that a person can have genuine knowledge of a good ethical quality if they are not

acting in a way that exhibits that quality. This is indeed in the vicinity of a denial of the possibility of *akrasia*. But it is not, strictly speaking, a denial of this possibility: Wang doesn't characterize the relevant knowledge as knowledge of a proposition (i.e. that an action is required of one). Moreover, he focuses on genuine knowledge, whereas whether a person is *akratic* depends on what they just plain know.

In fact, on the (standard) characterization of *akrasia* I just gave, it is natural to think that Wang would have held that *akrasia* is absolutely pervasive: according to him, any instance of voluntarily failing to do what one ought to do is an instance of *akrasia*. It is plausible that Wang held that *liangzhi* invariably produces inclinations which are appropriate to a person's circumstances; if there is an action a person ought to perform, their *liangzhi* would produce an inclination to perform it. Moreover, everyone – by the automatic, natural capacity of *liangzhi* – knows the ethical qualities of their inclinations by introspection. Wang says again and again that *liangzhi* knows these qualities regardless of the person's moral condition: if a person has an ethically wrong or bad response, *liangzhi* also knows its wrongness or badness. So according to Wang, anyone who fails to perform an action which they ought to perform would know the rightness or goodness of the inclination they failed to act on (as well as the badness of the inclination they acted on). It is a short step from this claim to the claim that such a person knows that the inclination they failed to act on was an inclination to perform an action that was required of them (and that they ought not have performed the action they did perform). If Wang took this short step, he would have have held that anyone who fails to perform an action which they ought to perform is *akratic*, since they act against the automatic knowledge of their *liangzhi*.

References

Angle, Stephen. 2018. Buddhism and Zhu Xi's Epistemology of Discernment. *Pages 156–192 of:* Makeham, John (ed), *The Buddhist Roots of Zhu Xi's Philosophical Thought*. Oxford University Press.

Angle, Stephen C. 2005. Sagely ease and moral perception. *Dao*, **5**(1), 31–55.

Angle, Stephen C. 2010. Wang Yangming as a Virtue Ethicist. *Pages 315–335 of: Dao Companion to Neo-Confucian Philosophy*. Springer.

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

Araki (荒木見悟), Kengo. 2017. 佛教與儒教. 國科會經典譯注計畫. trans. Liao Zhaoxiang 廖肇享.

Bol, Peter Kees. 2008. *Neo-Confucianism in history*. Cambridge, MA: Harvard University Asia Center.

Brennan, Tad. 2005. *The Stoic life: Emotions, duties, and fate*. Oxford University Press.

Chan, Wing-tsit. 1963. *Instructions for Practical Living and other Neo-Confucian Writings by Wang Yang-ming*. Columbia University Press.

Chan (陳榮捷), Wing-tsit. 1983. 王陽明傳習錄詳註集評. 臺灣學生書局.

Chang, Carsun. 1955. Wang Yang-ming's Philosophy. *Philosophy East and West*, **5**(1), 3–18.

Chen, Lai (陳來). 1991. 有無之境: 王陽明哲學的精神. 北京: 人民出版社.

Cheng, Chung-ying. 1974. Conscience, mind and individual in Chinese philosophy. *Journal of Chinese Philosophy*, **2**(1), 3–40.

Ching, Julia. 1972. *The philosophical letters of Wang Yang-ming*. Canberra: Australian National University Press.

Ching, Julia. 1976. *To Acquire Wisdom: The Way of Wang Yang-ming*. Columbia University Press.

Cua, A. S. 1993. Between Commitment and Realization: Wang Yang-Ming's Vision of the Universe as a Moral Community. *Philosophy East and West*, **43**(4), 611–647.

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

Fraser, Chris. 2011. Knowledge and error in early Chinese thought. *Dao*, **10**(2), 127–148.

Frisina, Warren G. 1989. Are Knowledge and Action Really One Thing?: A Study of Wang Yang-ming's Doctrine of Mind. *Philosophy East and West*, **39**(4), 419–447.

Fuss, Peter. 1964. Conscience. *Ethics*, **74**(2), 111–120.

Graham, Angus Charles. 1958. *Two Chinese Philosophers*. London: Lund Humphries.

Huang, Yong. 2008. Wang Yangming between Humeans and Anti-Humeans: Liangzhi as Besire (Belief/Desire) and not Bizarre 王陽明在休謨主義和反休謨主義之間: 良知作為體知 = 信念、欲望 ≠ 怪物. *Pages 147–65 of:* 陳少明, Chen Shaoming (ed), *Embodied Knowledge In Human Sciences* 體知與人文學. 北京: Huaxie Press 華夏出版社，.

Huang, Yong. 2017. Knowing-That, Knowing-How, or Knowing-To? *Journal of Philosophical Research*, **42**, 65–94.

Ivanhoe, Philip J. 1996. *Chinese Language, Thought, and Culture: Nivison and His Critics*. Open Court Publishing.

Ivanhoe, Philip J. 2000. *Confucian moral self cultivation*. Hackett Publishing.

Ivanhoe, Philip J. 2002. *Ethics in the Confucian tradition: The thought of Mengzi and Wang Yangming*. Hackett Publishing.

Ivanhoe, Philip J. 2009. *Readings from the Lu-Wang school of neo-confucianism*. Hackett Publishing.

Ivanhoe, Philip J. 2011. McDowell, W ang Yangming, and Mengzi's contributions to understanding moral perception. *Dao*, **10**(3), 273.

Ivanhoe, Philip J. 2018. *Oneness: East Asian Conceptions of Virtue, Happiness, and How We Are All Connected*. Oxford University Press.

Matilal, Bimal Krishna. 1986. *Perception: An essay on classical Indian theories of knowledge*. Oxford University Press.

Mou, Zongsan. 1973. The Immediate Successor of Wang Yang-ming: Wang Lung-hsi and his Theory of ssu-wu. *Philosophy East and West*, **23**(1/2), 103–120.

Mou, Zongsan (牟宗三). 1972. 王學的分化與發展. 新亞學術年刊, **14**, 89–131.

Nivison, David. 1996a. Reply to Ivanhoe. *In:* Ivanhoe, Philip J. (ed), *Chinese Language, Thought, and Culture: Nivison and His Critics*. Open Court Publishing.

Nivison, David S. 1973. Moral Decision in Wang Yang-ming: The Problem of Chinese "Existentialism". *Philosophy East and West*, **23**(1/2), 121–137.

Nivison, David S. 1996b. *The ways of Confucianism: Investigations in Chinese philosophy*. Open Court Publishing.

Peng (彭國翔), Guoxiang. 2003. 良知學的展開: 王龍溪與中晚明的陽明學. 臺灣學生書局.

Perrett, Roy W. 2016. *An Introduction to Indian Philosophy*. Cambridge University Press.

Peterson, Willard. 1986. Another Look at Li 理. *Bulletin of Sung and Yüan Studies*, 13–31.

Qian (錢明), Ming. 2005. 儒學 "意" 范疇與阳明學的 "主意" 話語. 中国哲學史, 11–18.

Shu, Jingnan (束景南), & Zha, Minghao (查明昊) (eds). 2016. *Wang Yangming quanji bubian* 王陽明全集補編. 上海: 上海古籍出版社.

Shun, Kwong-loi. 2010. Zhu Xi's Moral Psychology. *Pages 177–195 of: Dao Companion to Neo-Confucian Philosophy*. Springer.

Shun, Kwong-loi. 2011. Wang Yang-ming on Self-Cultivation in the *Daxue*. *Journal of Chinese Philosophy*, **38**(s1), 96–113.

Tang, Chun-i. 1970. The development of the concept of moral mind from Wang Yang-ming to Wang Chi. *Pages 93–119 of:* Theodore, William Theodore (ed), *Self and Society in Ming Thought*. New York: Columbia University Press.

Tang, Chun-i. 1973. The Criticisms of Wang Yang-ming's teachings as raised by his contemporaries. *Journal of Chinese Philosophy*, **23**(1/2).

Tien, David W. 2010. Metaphysics and the Basis of Morality in the Philosophy of Wang Yangming. *Pages 295–314 of: Dao Companion to Neo-Confucian Philosophy*. Springer.

Tien, David W. 2012. Oneness and Self-Centeredness in the Moral Psychology of Wang Yangming. *Journal of Religious Ethics*, **40**(1), 52–71.

Wu, Guang (吳光) (ed). 2007. *Collected Works of Liu Zongzhou* 劉宗周全集. 浙江: 浙江古籍出版社.

Wu, Guang (吳光), Qian, Ming (錢明), Dong, Ping (董平), & Yao, Yanfu (姚延福) (eds). 2011. *Collected Works of Wang Yangming* 王陽明全集. 上海: 上海古籍出版社.

Yong, Huang. 2015. *Why be Moral?* SUNY Press.

A Identity

There are a number of important arguments *against* attributing Identity to Wang. Later in the discussion with Xu Ai, parts of which appeared as **[T1]** and **[T2]** above, Wang says:

[T14] 某嘗說知是行的主意，行是知的功夫；知是行之始，行是知之成。
I have said that knowledge is the mastering aim (*yi* 意) of action and action is the practice of knowledge; that knowledge is the beginning of action and action is the completion of knowledge. (*IPL* 5, *QJ* 5 cf. *IPL* 26, *QJ* 15)[77]

Both the remarks which precede and those which follow this quotation discuss the unity of knowledge and action explicitly. Insofar as Wang restricted the doctrine of the

[77] There are two ways of taking the beginning of the passage, depending on how we understand *yi* 意 there. On a first interpretation (represented in my translation), 意 is taken as "aim" (for texts demonstrating that there is this meaning, see below n. 41). His idea is that one's action aims at knowledge, and acting is the process of fulfilling that aim. On this reading, Wang in fact emphasizes how knowledge and action are jointly achieved. On a second interpretation, *yi* means "inclination" and Wang speaks here of one or more mental events that last through the action and guides it. (So, Chen (1991, p. 101), and also Tiwald and Angle, who translate this "intent of acting".) This interpretation is a little in tension with the introspective model, since knowledge does not direct the action on that picture. But the tension is slight; even given the introspective model one can understand how Wang would (also) think of the knowledge in question as directing the action – it is similar to the way that homeostatic equilibrium guides the body's staying in a certain state. (I don't think we should see the "knowledge" here as describing a faculty, given Wang's focus on the unity of knowledge and action in the passage.)

In any case, I think there are reasons to prefer the first interpretation. One (not too strong) reason is that (as Chen (1991, p. 101) points out) if we adopt the second reading it is not clear why Wang's doctrine of the unity of knowledge and action is revolutionary at all, since his remarks would seem to suggest a picture on which knowledge comes first and action comes later, exactly the view he criticized his contemporaries for holding. A second much stronger reason is that the contrast between "guiding aim" 主意 and "practice" 工夫 was a formulaic one, that Wang uses at least twice in contexts which do not appear to have anything to do with a particular person's psychological state (*IPL* 25, *QJ* 15; *IPL* 168, *QJ* 80-1).

unity of knowing and acting to genuine knowledge, and this discussion concerns that doctrine, it is natural to see him here as focused on genuine knowledge, although he does not make that qualification explicit. As many authors have noted, passages like this are on their face inconsistent with the identity of (genuine) knowledge and action.[78] Knowledge and action have different properties; one is the beginning, and the other is the completion.[79]

There is a general conceptual reason to reject Identity, too. That principle entails that every episode of genuinely knowing Fness is identical to an event of acting Fly, and that every instance of acting Fly is identical to an instance of genuinely knowing Fness. This last claim is implausible, even given Wang's background commitments. The sage-king Wu raised an army without mourning. His actions of sending messengers to rouse the troops, of riding from place to place with his generals, or of marching with the army were not identical to anything understood to be knowing in any ordinary sense of that term. One can understand a conception of action on which his actions had various mental states or activities as *parts* (and also on which the mental events themselves were actions, see n. 45), but surely his actions involved a great deal more than this: he had to move his mouth, leap onto his horse or march.

This kind of argument can have at best modest evidential import for determining what Wang believed. The fact that a doctrine is implausible is not conclusive grounds for denying that Wang held it.[80] But I do think it means that the textual evidence would have to be strongly in favor of Identity if we were to attribute it to him.

The strongest textual evidence in favor of Identity comes from a family of passages in which Wang repeats the same formula. For ease of exposition, I present two translations of the formula (citations of its occurrences will be given below):

[T15] 知之真切篤實處，即是行；行之明覺精察處，即是知。

(A) Insofar as knowledge is genuine, practical, earnest and substantial, it is action; insofar as action is lucidly aware and precisely discriminating, it is knowledge.

(B) The genuineness, earnestness, practicality and substantialness of knowledge is action; the lucid awareness and precise discrimination of action is knowledge.

The difficulty in translating this passage derives from Wang's use of the expression *zhi chu* 之處. Here, I'm going to consider a range of interpretations, but tentatively

[78] See Chen (1991, p. 99), Araki (荒木見悟) (2017, p. 281-3)

[79] One might think this passage shows that Wang did not hold KA and AK: if knowledge is the beginning it could come before acting, so that there would be a time when one knows but isn't acting. But we could understand "beginning" to mean "first part" here, and then Wang's thought is that knowing is the first part of (bodily) action, and (bodily) action is the completion of the action of which knowledge is the first part.

[80] And indeed, some scholars have responded to the conceptual point by suggesting that Wang did indeed endorse a highly non-standard conception of knowledge, where (essentially by stipulation) knowledge *is* identical to these bodily actions. I believe this position is something we should retreat to only as a last resort.

suggest that we should prefer the (B) translation.

What actions are lucidly aware and precisely discriminating? The interpretation (A) yields different verdicts on whether Identity is true, depending on how we answer this question.

On option (A1), Wang held that only mental actions are lucidly aware and precisely discriminating. This position is best motivated if Wang endorsed Mental Actions (see n. 45). If we focus on inclinations as stereotypical mental actions (and perhaps as the most generic form of mental action), then the second part of the slogan above could be read as "insofar as an inclination is lucidly aware and precisely discriminating, it is knowledge".[81] This general line of interpretation derives support from an array of other passages where Wang makes similar claims about the connections between inclinations and knowledge, for instance:

[T16] 故無心則無身，無身則無心。但指其充塞處言之謂之身，指其主宰處言之謂之心，指心之發動處謂之意，指意之靈明處謂之知，指意之涉著處謂之物：只是一件。

So if there is no mind, there will be no body, and if there is no body, there will be no mind. Insofar as it fills space, it is called the body. Insofar as it is the master, it is called the mind. Insofar as the mind is aroused and set in motion, it is called an inclination.[82] Insofar as the inclination is lively and lucid, it is called knowledge (*zhi*). Insofar as the inclination is attached it is called a thing. They are all one piece. (*IPL* 201, *QJ* 103)

This passage uses the same controversial expression *zhi chu* 之處 to make the claim that inclinations, insofar as they are lively and lucid, are knowledge.[83] Wang's use of the expression "lively and lucid" here supports the idea that Wang held that only mental actions are lively and lucid, and thus that this is his focus in the present passage.

[81] On this interpretation the second part of the slogan supports Instance KA, see n. 65.

[82] The word *fa* has both a transitive and an intransitive use. In the intransitive use it means "is aroused". (Wang uses the word in this way in e.g. *IPL* 226, *QJ*, 109-110; it is also probably the best interpretation of arguably the single most important canonical discussion of emotions (in the *Doctrine of the Mean*), where the contrast is drawn between a situation in which they are "not yet aroused" (未發) and one in which they are "already aroused" (已發).) In its transitive use, the word means "emit". One might argue that the cases which I translate "arouse" would be better rendered as "emit" with an implied object, but the coupling with 動 which is not transitive here tells against this.

[83] There is thus an alternative interpretation of this passage, which is an analogue of the (B) interpretation above. On this interpretation we might translate the key sentences as: "The mind's arousal and moving is an inclination. The liveliness and lucidity of an inclination is knowledge." On this interpretation, the inclinations are identified with a property instance of arousal, or with the event of the mind being aroused; knowledge is identified with a property instance of liveliness and lucidity or the event of the inclination being lively and lucid. Whereas on the interpretation in the main text, the inclination (insofar as it is in a given condition) is knowledge, on this alternative, a given condition of the inclination is knowledge. I'll return to this alternative in subsequent notes; the difference doesn't affect the main points in the main text.

Whether we adopt this construal or the one in the main text, the idea behind saying that these are "all one piece is the same. In *IPL* 38 (*QJ* 18), Wang analogizes related ideas to the way "a person is one: in regard to his father he is called a son, and in regard to his son he is called a father" (人一而已：對父謂之子，對子謂之父).

This interpretation would not support Identity. For Wang would only say that episodes of genuine knowledge count as actions, and that appropriate mental actions count as genuine knowledge. These claims do not bear on the question whether all virtuous action is (genuine) knowledge.[84]

On option (A2) Wang held that both mental and bodily actions can count as lucidly aware and precisely discriminating. On this reading, the passage supports Identity. For presumably Wang would have held that actions had these qualities whenever they are virtuous. So in the passage, Wang would say precisely that episodes of genuine knowledge are actions, and (what is more surprising) that virtuous actions are episodes of knowledge.

Finally, on the (B) translation, Wang's idea would be something we might more fully express as "knowledge's being genuine, practical, earnest and substantial comes from action; action's being lucidly aware and precisely discriminating comes from knowledge". Wang follows the remark by emphasizing that there is just one effort of cultivating action and knowledge. He goes on to explain that this is because genuine knowledge is the form of knowledge that facilitates action, but at the same time, it wouldn't be genuine if it weren't followed by action. Wang's point is that what makes one's knowledge genuine, practical, earnest, and substantial is the action that issues from it; conversely what makes one's action lucidly aware and precisely discriminating is the knowledge it issues from. On this translation the passage does not support Identity.

The different passages in which Wang repeats this slogan give us a richer picture of his understanding of it. In these passages I'll print the (A) translation, though of course both are possible:

[T17] 知之真切篤實處，便是行；行之明覺精察處，便是知。若知時，其心不能真切篤實，則其知便不能明覺精察；不是知之時只要明覺精察，更不要真切篤實也。行之時，其心不能明覺精察，則其行便不能真切篤實；不是行之時只要真切篤實，更不要明覺精察也。

Insofar as knowledge is genuine, practical, earnest and substantial, it is action; insofar as action is lucidly aware and precisely discriminating, it is knowledge. If when you are [engaged in] knowing, your mind is unable to be genuine, practical, earnest and substantial, then your knowing will not be able to be lucidly aware and precisely discriminating; it is not that when you are [engaged in] knowing you only need to be lucidly aware and precisely discriminating, but don't need to be genuine, practical, earnest and substantial.

[84] On this interpretation, the first part of the slogan would make most sense if Wang's point was that not all knowledge-episodes, but *only* episodes of genuine knowledge count as action (i.e. if he endorsed Total Action; see n. 45). But we could still make sense of it otherwise; his point might be that knowledge insofar as it is genuine, practical, earnest and substantial *leads to bodily* action. He would then switch topics to describe mental action in the second part of the slogan.

This interpretation of the slogan lends some support to Simple Knowledge or Total Knowledge, as opposed to Introspective Knowledge. For it is natural to read **[T16]** as allowing that inclinations other than those which are *liangzhi*'s knowing of the goodness or rightness of an inclination can be lively and lucid. If those too can be genuine knowledge, then it must be that not all genuine knowledge is the knowledge of *liangzhi*, as I am committed to holding.

If when you are acting, your mind is unable to be lucidly aware and precisely discriminating, then your action will not be able to be genuine, practical, earnest and substantial; it's not that when you are acting, you only need to be genuine, practical, earnest and substantial, but don't need to be lucidly aware and precisely discriminating. (*QJ* 234)

It was more standard to speak of earnest *action* (篤行) (not earnest knowledge) and luminous or discriminating *knowledge* (not luminous or discriminating action). Wang's point in the passage is that knowledge can't have its typically valedictory properties unless it is associated with action which has *its* typical valedictory properties. The context of this passage seems to suggest the interpretation I offered for the (B) translation: the idea isn't to make a metaphysical claim about the nature of these actions or episodes of knowledge (as both of the (A) interpretations have it), but to explain how it is that knowledge (or action) achieves properties that make it the highest form of knowledge (resp. action). Reversing the ascription of properties helps to make that point, since it illustrates that knowledge in a certain sense depends on action (and vice versa).

In a section of the important letter to Gu Dongqiao, Wang quotes Gu as bringing up something closely related to our present concern – can Wang really mean that knowledge and action are identical? Wang's reply is illustrative:

[T18] 來書云：「『真知即所以為行，不行不足謂之知』，此為學者喫緊立教，俾務躬行則可。若真謂行即是知，恐其專求本心，遂遺物理，必有闇而不達之處。抑豈聖門知行并進之成法哉？」

知之真切篤實處，即是行；行之明覺精察處，即是知，知行工夫本不可離。只為後世學者分作兩截用功，失却知行本體，故有合一併進之說。真知即所以為行，不行不足謂之知，即如來書所云「知食乃食」等說可見，前已略言之矣。此雖喫緊救弊而發，然知行之體本來如是，非以己意抑揚其間，姑為是說以苟一時之效者也。

Your [Gu's] letter says: " 'Genuine knowledge is just what is used to promote action; if you don't act it is not enough to be called knowledge'. This is acceptable as an urgent teaching for students, to get them to devote themselves to personal action. But if you truly mean that action just is knowledge, I'm afraid that they will exclusively seek the original mind and straightaway leave behind the *li* of things, so that there must be some places where they are blocked and which they cannot fully penetrate. Is this really the sages' established method of the joint promotion of knowledge and action?"

[Wang's reply:] Insofar as knowledge is genuine, practical, earnest and substantial, it is action; insofar as action is lucidly aware and precisely discriminating, it is knowledge. The practice of knowledge and action at root cannot be separated. It is only because later generations of students have divided them into two separate practices and lost the original substance of knowledge and action that I have proposed the theory of their unity and joint promotion. Genuine knowledge is what is used to promote action; unless one acts it is not enough to be called knowledge – this is just what can be seen from the exam-

> ple of "knowing the food and then eating" in your letter, which I've discussed briefly above already. Although this is something I put forward to encourage people and to rescue them from a fault, the substance of knowledge and action is originally like this – it's not that I followed my own inclinations and promoted or demoted one of them, carelessly proposing this theory as expedient for the present time. (*IPL* 133, *QJ* 47-8)

Some of what Wang says here might seem to point toward Identity. Wang does not directly deny that he believes the claims that Gu attributes to him. Instead, it might seem that he defends those claims. But on inspection, this isn't quite what happens in the passage. For Wang fairly clearly takes Gu's charge to be focused on the doctrine that the *practice* of knowledge and action is the same, and that they are *jointly promoted*. Indeed, just as Gu describes his slogan as "the joint promotion of knowledge and action", Wang speaks of "unity and joint promotion" as a unit in his response. The defense Wang gives of his position is not a defense against the kind of conceptual problem with Identity sketched a few paragraphs ago. Rather it is a defense of the idea that one must cultivate knowledge and action together, and that "exclusively seeking one's original mind" will not in fact lead to the kind of self-absorbed separation from worldly affairs Gu feared it would. Perhaps clearest of all, the discussion leads with (and returns to) a claim which is incompatible with Identity, the claim that genuine knowledge is what is used to promote action. Like the claim that knowledge is the beginning of action, this claim ascribes to knowledge different properties than it ascribes to action, and so rules out the true Identity of the two.[85]

A number of other related passages support the idea that, to the extent Wang makes remarks in the vicinity of Identity, he means to be advancing a principleabout the *practice* (工夫) of knowledge and action (i.e. how one achieves the highest forms of knowledge and action), not about the metaphysics of the two. In yet another passage where Wang uses the slogan, he emphasizes directly that he means merely that the practice or method of cultivating knowledge and action are identical (元來只是一個工夫, *QJ* 232).[86] Although not directly in connection to the slogan, some of his most emphatic claims about the intimate connection between knowledge and action concern the identity of this practice (e.g. 知行原是兩個字說一個工夫，這一個工夫須著此兩個字，方說得完全無弊病 *QJ* 233, cf. *IPL* 5, *QJ* 5). And, as in **[T18]**, later in the letter to Gu Dongqiao, Wang again renders "the unity of knowledge and action" (知行合一) more fully as "the unity and joint promotion of knowledge and action" (知行合一并進), concluding a long discussion of how apparent examples of study in fact require action: "So you know that the knowledge and action are unified and jointly promoted, and that they cannot be sep-

[85] This argument isn't quite as conclusive as I would like. If Wang had written 真知即所以行, this would uncontroversially mean that genuine knowledge is used in guiding action; it is what is used to act. But what he did write, 真知即所以為行 has three different possible construals. If we read *wèi* here, it means (i) that genuine knowledge is used to promote action. If we read *wéi* it most likely means (ii) that genuine knowledge is used to encourage action. But it is perhaps *just* possible that the expression linguistically could mean (iii) that genuine knowledge constitutes action. Still, this interpretation is much less natural than the alternatives, so on balance I think the passage does point away from Identity.

[86] Chen (1991, p. 103) also emphasizes this point.

arated into two separate tasks (*shi*)" (則知知行之合一并進，而不可以分為兩節事矣。 *IPL* 136, *QJ* 52). These repeated remarks suggest that Wang might even have seen "the unity of knowledge and action" as an abbreviation for "the unity and joint promotion of knowledge and action", a fuller version which does not suggest Identity.

The (A) translation of **[T15]** strikes me as the most natural local linguistic option. So, given the arguments against Identity, if one were moved by linguistic evidence alone, I'd adopt (A1). But while I concede the linguistic construal of (B) (and how it relates to the associated interpretation) is on its own less natural, it fits better with what Wang says in the array of other places I've just discussed.[87] Wang's point is that one must cultivate knowledge and action together.

In any case, these passages do not provide clear support for Identity; they certainly do not provide evidence of the kind of emphatic insistence on Identity that would be needed to ascribe such an implausible doctrine to Wang. My suspicion is that readers of Wang have been led to principles like Identity simply by overemphasis on a particular meaning of "unity" *heyi* 合一. But as I've said there's fairly direct evidence that Wang himself didn't understand his slogan as describing this kind of unity. The joint promotion of knowledge and action is an idea that fits well with KA and AK, but does not require Identity.

[87] Wang himself didn't take the meaning of the slogan to be obvious, as he says in a letter to Zhou Daotong 周道通, *QJ* 1331.