

The Introspective Model of Genuine Knowledge and the Unity of Knowledge and Action

Harvey Lederman

August 8, 2020

Please do not cite without permission, but comments welcome!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). Wang held that action was not unified with all knowledge, but only with an elevated form of knowledge, which he sometimes called "genuine knowledge". My interpretation focuses on a new understanding of this notion, according to which a person has genuine knowledge if and only if they are free from a particular form of doxastic conflict. I argue that Wang believed that a person is free from this form of doxastic conflict if and only if they are free from a form of motivational conflict. I suggest that Wang held moreover that a person is free from this form of motivational conflict if and only if they are acting virtuously. Thus, my interpretation helps to explain how Wang could hold that a person has genuine knowledge if and only if they are acting virtuously.

1 Introduction

In 1508, 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. In the following year, Wang distilled this dramatic revelation in the doctrine of the "unity of knowledge and action" (*zhi xing he yi* 知行合一).¹ The

¹I follow tradition in translating *he yi* 合一 as "unity" in this slogan. The expression can also mean something weaker, more like "correspondence". The account of Wang's "enlightenment" can be found in Wang's *nian pu*; see Wu *et al.* (2011, 1354-5).

Throughout the paper, I will cite passages from the *Instructions for Practical Living* (hereafter, *IPL*, 傳習錄) by the section number of Chan's editions (Chan (1963), Chan (陳榮捷) (1983)), followed by a page number in Wu *et al.* (2011) (indicated by "QJ"). Passages in Wang's works outside the *IPL* are cited by the *juan* number, a period, and then the page number (e.g. "QJ 6.242"). Where available, I also cite pages in the translations of Ching (1972) and Ivanhoe (2009). Most of these latter translations are reprinted, but

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, alongside Confucius, Mencius, and the great Song dynasty philosopher Zhu Xi (朱熹, Yuanhui 元晦, 1130-1200) (“孔孟朱王”).

But while the unity of knowledge and action has been widely celebrated, it has also been widely condemned. In letters and recorded conversations, Wang’s interlocutors presented a series of putative counterexamples to what they understood by his slogan. In his replies, Wang retreated to the claim that action is unified only with an elevated form of knowledge, which he sometimes called “genuine knowledge”. These replies suggest that Wang meant to stipulate that “genuine knowledge” should be understood as “whatever is unified with action”. Such a stipulation might make Wang’s doctrine true, but it would do so only by rendering it trivial. And thus, the “unity of knowledge and action” has come to seem at best a misleading advertisement for a triviality.

I will argue that this criticism is mistaken. I will develop a new interpretation of Wang’s views about genuine knowledge, according to which 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 argues that a person is free from this form of doxastic conflict if and only if they are free from a particular form of motivational conflict. He further connects freedom from this form of motivational conflict to virtuous action. The result is that Wang did not stipulate the truth of his doctrine. Instead, under the heading of the unity of knowledge and action, he advanced striking, substantive claims connecting freedom from doxastic conflict to virtuous action.

The paper is structured as follows. Section 2 frames the paper and introduces some key background. Section 3 examines Wang’s views about the conscience-like faculty of *liangzhi*. I argue that *liangzhi* invariably acquires knowledge that is importantly related to genuine knowledge, but this knowledge does not always amount to genuine knowledge. Section 4 discusses a previously unidentified argument in Wang’s writings. This argument, which I call the *obscuration argument*, describes what more is required, beyond *liangzhi*’s knowledge, for a person to have an elevated form of knowledge. Section

with some significant changes, in [Tiwald & Van Norden \(2014\)](#). I recommend that the reader consult the amended versions where possible, but since they do not cover all of the material translated in [Ivanhoe \(2009\)](#), I cite page numbers in the earlier book here.

5 completes the presentation of the *introspective model of genuine knowledge*, and offers my response to the main challenge of the paper. I show how Wang conceives of genuine knowledge as a form of knowledge, and how it is elevated above ordinary knowledge in a distinctively epistemic or doxastic respect. Section 6 concludes.²

2 Genuine knowledge and wholehearted inclinations

The aim of this paper is to present an account of genuine knowledge which elucidates the sense in which genuine knowledge is an elevated form of knowledge. But Wang introduces the notion of genuine knowledge in the course of responding to apparent counterexamples to his doctrine of the unity of knowledge and action; his views about the notion of genuine knowledge are inextricably bound up with his broader doctrine of the unity of knowledge and action. If an account of genuine knowledge did not fit with reasonable interpretations of this doctrine, that would be a strong reason to reject

²For reasons of space, I have been unable to include thematic discussion of others' interpretations of the unity of knowledge and action in this paper. A companion paper, [Lederman \(2020a\)](#), 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.

A word about chronology and the scope of this paper: Wang first advocated the unity of knowledge and action in 1509, a year after his "enlightenment" in Longchang. Wang's articulation of his views underwent an important shift twelve years later, in 1521, when he first began to emphasize the importance of *liangzhi* (see Qian Dehong's (錢德洪 Hongfu 洪甫, 1496-1574) account of this year in the *nian pu* (QJ 34.141) along with his famous account of the "three turns" (QJ 41.1745-6), cf. e.g. [Tu \(1976, p. 10-11\)](#), [Ching \(1976, p. 41-46\)](#)). But Wang continued to endorse the unity of knowledge and action explicitly after 1521, and at least as late as 1526; moreover, there is no evidence he retracted it before his death in 1529. (The latest explicit mentions I'm aware of are in "Reply to Inquiries from a Friend" (1526) QJ 6.232; [Ching \(1972, 106-8\)](#) and [Ivanhoe \(2009, p. 123-7\)](#). For other post-1521 discussion, see *IPL* 133 QJ 48, *IPL* 139 QJ 56 ([Chan \(1963, p. 91 n. 1\)](#) discusses difficulties with dating this latter); "Second Letter to Lu Yuanjing" (1522) QJ 5.210 [Ching \(1972, p. 68-9\)](#); *Letter to Zhu Yangbo* (1524) QJ 8.309.)

My aim in this paper will be to explicate the views Wang held about the unity of knowledge and action after 1521. In developing this interpretation, I will freely take conversations and writings prior to 1521 as evidence for Wang's views after this date. This practice is justified by my belief that, although Wang had not fully developed his later ideas before 1521, his early views were at least consistent with those he would come to hold later. There is strong evidence that Wang himself understood his philosophical development in this way. First, Wang did not retract or seek to revise a collection of his sayings published in 1518; nor did he intervene when that same collection was republished together with a number of letters in 1524 (on the publication history of the *IPL* see, [Chan \(1963, p. 314\)](#), [Ivanhoe \(2002, Appendix I\)](#); for the later publication history of the complete works, see [Chu \(1988\)](#)). Second, Wang is recorded as saying: "From Longchang on [the site of his "enlightenment"], I have not departed from the meaning of the two characters '*liangzhi*'. But I was just unable to produce these two characters in speaking to students, 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 41.1747) (All translations in the paper are mine, although I have always consulted [Chan \(1963\)](#), [Ching \(1972\)](#) and [Ivanhoe \(2009\)](#) for passages translated in those works.)

it. Conversely, if such an account helps us to understand the doctrine, that is some evidence in its favor. For these reasons, I will spend considerable effort in the paper showing how my account of genuine knowledge fits into an interpretation of Wang's overall doctrine. In this section, I introduce my strategy for achieving this goal, and then present a few key ideas about knowledge and genuine knowledge that will be important for the rest of the paper.

I will argue in detail that my account of genuine knowledge helps us to understand a connection between genuine knowledge and a motivational state Wang describes using the technical term "wholehearted inclinations" or "making inclinations wholehearted" (*cheng yi* 誠意).³ As I will now explain, it is uncontroversial that Wang believes that a person's having wholehearted inclinations is closely connected to their acting virtuously. Given this claim, showing that my account of genuine knowledge helps us to understand the connection between genuine knowledge and wholehearted inclinations will suffice to show that the account fits with Wang's broader doctrines concerning the unity of knowledge and action.

The expression "making inclinations wholehearted" comes from the *Great Learning* (大學), one of the most important canonical texts for scholars working in Wang's tradition. Part of this text was understood to contain an enumeration of eight aspects of the development of an ethical state, four of which – including "making inclinations wholehearted" – were taken to be stages of or tasks in a person's training themselves to be virtuous. Everyone in Wang's tradition agreed that a person would be wholly virtuous if and only if they had successfully completed all four of these tasks. Wang himself held a further, distinctive view about the relationship among these four tasks; he held that a person would have completed one of them if and only if they had completed all of them.⁴ As a consequence, Wang held that having wholehearted inclinations is not just necessary for a person to be wholly virtuous (as essentially all who took the *Great Learning* to be a canonical text would have agreed), but also sufficient; he held that a person is fully virtuous if and only if their inclinations are wholehearted. Since it is clear that Wang held that there is an important connection between a person's being fully virtuous and their performing virtuous actions, he was committed to holding that there is a corresponding, important connection between wholehearted inclinations and virtuous action.

The details of how Wang understood wholehearted inclinations, and how he con-

³The traditional (although opaque) translation of *cheng* 誠 is "sincere". *Cheng yi* is variously rendered as "making the will sincere", "making thoughts sincere" or "making intentions sincere".

⁴*QJ* 26.1069-70, Chan (1963, 277). See Shun (2011, Section IV) and Ching (1976, p. 82-4) for discussion.

nected wholehearted inclinations to virtuous action are controversial. But controversy over these questions is largely independent of questions about the character of genuine knowledge. So, I will focus in this paper on how genuine knowledge is related to wholehearted inclinations. In particular, taking my cue from an important passage (*IPL* 5, *QJ* 4) where Wang discusses knowledge of filial piety (*xiao* 孝, hereafter “filiality”) in connection to the unity of knowledge and action, I will develop an interpretation which vindicates the following claim:

Genuine/Wholehearted A person genuinely knows filiality if and only if they have a wholehearted inclination to perform a filial action.

My *introspective model of genuine knowledge* will consist of two principles – Knowing Good and Introspective Knowledge – which, together with some natural background assumptions, entail Genuine/Wholehearted. Vindicating this connection, as I have said, ensures that the account of genuine knowledge fits with many reasonable interpretations of the unity of knowledge and action, without requiring me to engage here in contentious and difficult debates about the character of wholehearted inclinations and their exact connection to virtuous action.⁵

Wang seems to have endorsed the spirit of Genuine/Wholehearted not only for filiality but also other virtues he speaks of in connection to the unity of knowledge and action, most obviously, fraternal respect (*ti*, 悌, hereafter “respect”), conscientiousness (*zhong* 忠), humaneness (*ren* 仁), and compassion (*ce yin*) 惻隱).⁶ Below, when I display principles like Genuine/Wholehearted which discuss filiality explicitly, I intend also to endorse the obvious variants of them for these other virtues, and when I use names for principles like “Genuine/Wholehearted” I will sometimes mean the family of these

⁵It may be helpful to have a more concrete sense of how a fuller account of the unity of knowledge and action might go. Here is one way, that I find attractive. Suppose we take the key principle behind Wang’s slogan to be:

Unity A person genuinely knows filiality if and only if they are acting filially.

And suppose, moreover, that Wang endorses:

Inclination Action A person is acting filially if and only if they have a wholehearted inclination to perform a filial action.

These principles, together with any theory which vindicates Genuine/Wholehearted would offer a *prima facie* attractive interpretation of the unity of knowledge and action. I argue for attributing Unity to Wang in [Lederman \(2020b\)](#).

⁶For respect, see *IPL* 5, *QJ* 4. For conscientiousness and humaneness, see *IPL* 139, *QJ* 56. For compassion, see *IPL* 135, *QJ* 50-1. I follow tradition in translating 惻隱 as “compassion” but the term might be better rendered as “being pained by” or “unable to bear”; see e.g. [Shun \(2018, p. 90\)](#) for discussion.

Beyond these examples, it becomes less clear how far Wang intended his ideas to apply; I give some discussion below, n. 48.

principles, not just the one which describes filiality. But for readability, I'll continue to present principles only for filiality explicitly, and leave it to the interested reader to produce the relevant variants.

Officially, in this paper, I will not commit myself on how the notion of wholehearted inclinations is to be understood. But it will sometimes be helpful to have a concrete interpretation of this notion before us to make some of Wang's ideas vivid. On my favored interpretation, Wang holds that a given inclination of a person is wholehearted if and only if the person has no other inclinations which conflict with it. So, for instance, Wei's inclination to cool his parents in summer (a paradigmatic example of a filial action) is wholehearted if and only if he has no other inclination which conflicts with this one. If he has an inclination to go hiking in the mountains on a given day, leaving his parents sweltering behind him, then even if he stays to cool his parents, his inclination to cool them is not wholehearted. On this interpretation, to have a wholehearted inclination is to be free from motivational conflict of a particular form.

This completes the introduction of my strategy for showing how genuine knowledge fits into the unity of knowledge and action more broadly. Before turning to develop the account in earnest, I must introduce some important ideas which are raised to salience by Genuine/Wholehearted: first, about the word I am translating as "knowledge"; second, about what texts I take to be relevant to the interpretation of "genuine knowledge"; and, third, about the idea that genuine knowledge is a form of objectual knowledge ("knowing filiality").

First, when it occurs as a free-standing semantic unit, I will translate the character *zhi* 知 by "knowledge" and its cognates throughout this paper.⁷ If Wang had wanted to translate the English "Wei knows that Xin loves Yun", he would have used the character *zhi*: this character can mean "know". But the character has a broader semantic range than "know". It can be used to describe a change of state, and in this usage it is most naturally translated as "realize" or "recognize". *Zhi* can also be used without a complement, and in this usage it is naturally rendered as "is conscious", as in the sentence "Wei is conscious again, after months in a coma". In one passage, for example, Wang is asked whether people in a dreamless sleep still *zhi*, i.e. whether they are in any sense conscious (*IPL* 267, *QJ* 120).

⁷For detailed, helpful discussion of the meaning of "know" in classical, pre-Han (before 202 BCE) texts see Harbsmeier (1993). Geaney (2002) is an important treatment of the epistemology of sense perception in that period; Fraser (2011) argues for a particular conception of knowledge among the Mohists and in the *Xunzi* (both also pre-Han) and ties this conception to an alleged relative lack of discussion of skeptical arguments in this period. Angle & Tiwald (2017, Ch. 6) and now Angle (2018) argue that Zhu Xi (1130-1200) held there were at least three forms of knowledge, which he ranked in terms of cognitive achievement and practical importance.

Wang's theoretical remarks about *zhi* seem to pick up on these latter two uses of the character *zhi* ("recognize", conscious awareness) more than on its use to describe a persistent state of knowledge. In a battery of important passages, Wang describes the relationship between *zhi* and inclinations *yi* (意).⁸ In all of these passages, he describes *yi* as short-lived mental episodes, which "are aroused" or "are moved", not long-lasting states.⁹ In two of them, he characterizes inclinations (*yi* 意) as the "mind when it is aroused and moving" (心之發動), and then describes *zhi* as a property or aspect of such inclinations (IPL 174 QJ 86-7; IPL 201 QJ 103).¹⁰ When Wang says that *zhi* arises as a feature of the short-lived episodes of *yi*, he must be thinking of *zhi* itself as a short-lived episode. The passages strongly suggest that when Wang was attending closely to *zhi* he was sometimes interested in episodes of recognition, or perhaps more extended episodes of apprehension or grasp.¹¹ Although I'll continue to translate *zhi* as "know" throughout, in my explanations of various passages, I'll sometimes use the expression "episode of knowing" to indicate that I think Wang is focused on this special theoretical notion of *zhi*-ing. I don't intend this expression to be interpretable on its own; it is meant as a place-holder that will come to have significance in light of the way it appears in

⁸IPL 6 QJ 6; IPL 78 QJ 27; IPL 137 QJ 53; IPL 174 QJ 86-7; IPL 201 QJ 103.

⁹It is not clear to me that this aspect of Wang's theoretical view of *yi* corresponds with ordinary usage of the word in the literary Chinese of Wang's day, but it is clear that it differs from what the English word "inclination" describes. If Wei has an inclination to visit his family over the new year, he may have that inclination over a period of months or longer, and regardless of whether he is considering his plans or not at a given moment. Wang's theoretical remarks about *yi* suggest that for him, *yi* differ from inclinations in this respect: he seems to take them to be mental episodes, which arise and disappear fairly quickly, rather than longstanding states.

¹⁰In the other passages, Wang instead describes *zhi* as the "original intrinsic essence" of inclinations (*benti* 本體, IPL 6 QJ 6), takes inclinations to be "the arousal and moving" of *zhi* (發動 IPL 78, QJ 27) and says that inclinations are *liangzhi* when "it is stimulated, and is responsive and moving" (感應而動 IPL 137 QJ 53). These passages show that he could also use the character *zhi* to describe a capacity for producing the episodes I describe in the main text. For further discussion see nn. 11 and 19 below.

¹¹It is not clear to me that Wang always uses "genuine knowledge" and related expressions to describe episodes. Sometimes he seems instead to identify this elevated form of knowledge with a *disposition* to experience episodes of the relevant kind. On this way of using "genuine knowledge", Wang would endorse not Genuine/Wholehearted but:

Dispositional Genuine/Wholehearted A person experiences an episode of the exercise of genuine knowledge of filiality if and only if they have a wholehearted inclination to perform a filial action.

If forced to guess I would say that Wang had not settled on univocal senses for this family of technical terms: he could use "genuine knowledge" and related expressions to describe episodes of knowing but also to describe a disposition to experience such episodes. If this guess is right, then provided we disambiguate the term "genuine knowledge" Wang would have endorsed both Genuine/Wholehearted and Dispositional Genuine/Wholehearted. But the issues that will concern us in this paper most clearly concern episodes of the exercise of the relevant disposition, not the disposition itself. So, to simplify the presentation, I will focus on Genuine/Wholehearted throughout the paper. Nothing of substance hinges on this: everything I say would carry over straightforwardly, with relevant changes of terminology, using the alternative principle.

Wang's theory.¹²

Second, my term “genuine knowledge” is a translation of an expression composed of two characters, the first of which can be translated as “genuine”, “real”, “true”, or “authentic” (*zhen* 真), and the second of which is the character *zhi* we have just been discussing. Wang uses this expression to describe an elevated form of knowledge connected to the unity of knowledge and action in a number of passages.¹³ But Wang also uses an expression I will often translate by “extended knowledge” (*zhizhi* 致知) – which, alongside wholehearted inclinations, is one of the four aspects of personal ethical training described in the *Great Learning* – in a closely related way, to describe an elevated form of knowledge connected to the unity of knowledge and action.¹⁴ Wang can use “extended knowledge” to describe mental states or events that are not relevant to the unity of knowledge; “extended knowledge” can be used where “genuine knowledge” cannot be. But it is natural to suppose that in passages where Wang does connect extended knowledge to the unity of knowledge and action he is describing the same elevated form of knowledge that he describes with “genuine knowledge”.¹⁵ Thus, while

¹²It 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 epistemology in Europe. 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).

¹³*IPL* 5, *QJ* 4; *IPL* 5, *QJ* 5 (知得真); *IPL* 125, *QJ* 42 (where it is used by a disciple, not by Wang himself); *IPL* 133, *QJ* 47-8. (The occurrences of the expression in *IPL* 134, *QJ* 49 and *QJ* 20.829 are not related to the unity of knowledge and action.) In two different texts Wang ties what seems to be the set phrase “Insofar as knowledge is genuine, practical, earnest and substantial, it is action; insofar as action is lucidly aware and precisely discriminating, it is knowledge” (知之真切篤實處，便是行；行之明覺精察處，便是知。) to the unity of knowledge and action (*IPL* 133, *QJ* 48; *QJ* 6.233, 234; Ching (1972, 106-8)). In a few places, Wang also speaks of an advanced stage of ethical training being associated with *liangzhi* itself being “genuine and practical” (真切) (*IPL* 241 *QJ* 114; *QJ* 6.238). In these passages, Wang describes a property of the faculty of *liangzhi*, not of the episodes of knowledge produced by that faculty, but it is natural to think that such a property of the faculty would carry over to the episodes produced by it. In *IPL* 170 *QJ* 83, Wang connects the achievement of this genuineness and practicality to the extension of *liangzhi*, a notion he elsewhere (see next note) ties to the unity of knowledge and action.

¹⁴*IPL* 139 *QJ* 56; *IPL* 140, *QJ* 58; *IPL* 321, *QJ* 137; *QJ* 5.211, Ching (1972, pp. 68-9); *QJ* 8.308; and *QJ* 27.1100. A rougher connection is drawn in: *QJ* 6.234, Ching (1972, 106-8).

¹⁵In some places Wang uses the expression “the extension of knowledge” generally to describe cultivating any capacity of *liangzhi*. But these capacities are quite broad (see n. 29 below, for discussion). Perhaps most notably, in *IPL* 171 (*QJ* 83-4) Wang talks about extending *liangzhi* in connection to anticipating others’ actions (i.e. knowing what they will do in the future) on the basis of their present intentions. The criterion just described down does not imply that the “extended knowledge” described in this passage is genuine knowledge, and indeed it is natural to suppose that it is not, in part because it is unclear how a principle like Genuine/Wholehearted could apply to this case.

I will use “genuine knowledge” throughout as my technical term for this elevated form of knowledge, I take uses of “extended knowledge” which are connected to the unity of knowledge and action to describe genuine knowledge as well.

Third, and finally, I will now argue that some key uses of genuine knowledge should be understood to describe objectual knowledge, as I have formalized this idea in the principle Genuine/Wholehearted (“know filiality”). The word *zhi*, like the English word “know”, can describe *prima facie* different kinds of knowledge depending on the syntactic type of its complement: if it takes a sentential complement, it describes propositional knowledge (“know that”);¹⁶ if it takes a simple noun-phrase as its complement (“arithmetic”), it describes objectual knowledge (“know arithmetic”); and, finally, if it takes a verb-phrase as its complement (“run”, “say thank you”), it describes knowing-how (“know how to run”), or knowing-to (“know to say thank you”). In perhaps the most famous discussion of the unity of knowledge and action (*IPL* 5, *QJ* 4), Wang discusses what I have translated as objectual knowledge, “knowledge of filiality” and “knowledge of respect”. But this translation is controversial. The characters for “filial” and “respectful” can be used either as abstract nouns (e.g. “filiality”) or as verb-phrases (e.g. “be filial”), so the grammatical rules of classical Chinese do not on their own decide whether Wang here describes objectual knowledge (“know filiality”), knowledge-how (“know how to be filial”); or knowledge-to (“know to be filial”).¹⁷

But while the grammar on its own does not decide this issue, the broader context tells clearly in favor of objectual knowledge. In the passage, Wang presents seven examples, in three groups, to illustrate the way in which genuine knowledge is connected to action: knowledge of sights (*se* 色) and odors (*chou* 臭); knowledge of filiality and respect; and knowledge of pain (*tong* 痛), cold (*han* 寒) and hunger (*ji* 饑). There are strong arguments that in the first and third groups of examples, Wang describes objectual knowledge. This point is clearest for the first group (sight and odor), where the only grammatical possibility is that the relevant expressions describe objectual knowledge (“see a lovely color” 見好色, “smell a hateful odor” 聞惡臭 and “know the odor” 知臭). For the third group of examples (pain, cold and hunger) the argument

¹⁶A common way of expressing propositional knowledge in classical Chinese involves a special construction where a nominalized sentence is the complement of the verb “know”. At least in Wang’s corpus, there does not seem to me a significant semantic difference between this construction and those in which the complement is an un-nominalized sentence. See again Harbsmeier (1993) for discussion of a related issue in pre-Han texts.

¹⁷It is in some sense linguistically possible that Wang means to describe propositional knowledge (“know that one is filial”) but that construal makes little sense in context and I know of no one who has defended such a construal, so I won’t discuss it further here.

For favorable discussion of the knowing-how construal, see Huang (2008), and Ivanhoe (2000, p. 71, n. 15). For favorable discussion of the knowing-to construal, see Huang (2017).

is more involved, but still conclusive. In these cases, a construal as “knowledge-how” or “knowledge-to” is not ungrammatical: like the character for “filial”, the characters for “pain”, “cold” and “hunger” can also be used as stand-alone verb-phrases. But in the passage, Wang says that a person can know pain, cold and hunger, only by being pained, cold or hungry. And this idea is hard to make sense of on either the knowledge-how or the knowledge-to construals. For knowledge-how: to the extent that I can understand what it would mean to say that someone knows how to be cold in the relevant sense, surely what is surprising is that infants know how to be cold before they have ever been cold; it would seem obviously false to say that people know how to be cold only by being cold. For knowledge-to: again, to the extent that I can understand what it would mean to say that someone knows to be cold in the relevant sense (perhaps as “they know when to feel cold”), surely what is surprising is that infants know (when) to be cold, before they have ever been cold; it would seem obviously false to say that they know (when) to be cold only by being cold. By contrast, Wang’s idea is straightforward and intuitive on the objectual construal. On this construal his point would be that one does not know cold, pain or hunger, in the sense of being intimately acquainted with these states, until one has experienced them.

Since the objectual construal is the only construal that makes sense of the first and third groups of examples Wang uses to illustrate features of genuine knowledge, and since we should prefer an interpretation which gives a uniform sense to all seven of the examples, we should take Wang to be describing objectual knowledge of filiality and respect in the passage as well. Moreover, given that Wang describes genuine knowledge in this way in this important passage, it is natural to think that he takes genuine knowledge in general to be objectual knowledge.

These remarks allow us to refine the challenge presented in the introduction. Our aim will not just be to understand the sense in which genuine knowledge is an elevated form of knowledge, but moreover to understand the sense in which genuine knowledge is an elevated form of objectual knowledge of virtues like filiality.

3 *Liangzhi* and knowledge of ethical qualities

Wang’s most influential contributions to moral psychology concern the conscience-like faculty of *liangzhi*. In this section, I argue that Wang holds that a person has a good inclination if and only if their *liangzhi* knows that it is good. The fact that Wang believes that people have this knowledge of their own minds will set the stage for my examination of the connection between genuine knowledge and wholehearted inclinations,

which begins in the next section.

The word *liangzhi* 良知, which is made up of two characters, “pure” *liang* 良 and “knowledge” *zhi* 知, occurs in *Mencius* 7A.15, where it is connected to “pure ability” (*liangneng* 良能). In the *Mencius*, these two terms describe people’s putatively innate recognition that they should love their parents and respect their older brothers. Wang’s Song dynasty predecessors often discussed this passage of *Mencius*, but they tended to use the expression *liangzhi* together with *liangneng* and to cite the passage in general discussions of humans’ putative innate capacity for virtuous action.¹⁸

Wang split *liangzhi* from *liangneng* and developed a new multifaceted understanding of this notion, taken on its own. A central aspect of this new understanding was the view that a person’s *liangzhi* recognizes the ethical qualities of the person’s mental events. Thus Wang writes:

- [T1] 意則有是有非，能知得意之是與非者，則謂之良知。
Some inclinations are then right, and others wrong; what is able to know that inclinations are right and wrong is called *liangzhi*. (QJ 6.242; Ching (1972, p. 114))

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 and throughout the paper since Wang clearly thinks of this correctness/incorrectness as ethical or moral correctness/incorrectness.

In [T1], Wang describes *liangzhi* as a capacity; he says that it is able to know that inclinations are right and wrong.¹⁹ In other passages, however, he says not only that *liangzhi* can know that mental episodes are right or wrong, but that it does:

- [T2] 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 that they are right if they are right, and that they are wrong 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

¹⁸For a compact discussion of Wang’s predecessors, see Peng (彭國翔) (2003, 30-37).

¹⁹Throughout this paper, I refer to *liangzhi* as a “faculty”. But it is in fact controversial whether it is appropriate to speak of it in this way. On an alternative model of *liangzhi*, which I call the *Activity Model*, *liangzhi* is not a faculty but a set of events of awareness, emotions, inclinations and so on (Angle & Tiwald (2017, p. 104) describe it as a “category” of emotions). Proponents of this model must give an alternative interpretation of Wang’s talk of *liangzhi* as “what is able to” in passages like this one. My own view is that a *Faculty Model* of *liangzhi* is preferable, and I speak of *liangzhi* as a faculty for simplicity throughout, though I note that the main work of this paper can be done regardless of which model one adopts.

not yet encountered, but which I would render “concern” *nian*.²⁰ For the purposes of this paper, and to simplify the exegesis below, I will assume that Wang held that something is a motivating concern if and only if it is an inclination; the reader should thus treat these terms as interchangeable.²¹

[T2] says that *liangzhi* knows that a motivating concern is right or wrong whenever the motivating concern “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. Thus, the passage suggests that Wang held that *liangzhi* is not just able to know that motivation concerns are right or wrong, but that it invariably knows that they are right or wrong if they are.

In addition to knowing that inclinations are right and wrong, Wang also holds that *liangzhi* knows that inclinations are good (*shan* 善) and bad (*e* 惡), if they are.²²

[T3] 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。

²⁰This term is often rendered simply “thoughts”. But this translation does not capture the fact that *nian* have more of an affective (and even an action-directed) component than other mental events; for instance, they are more affect-laden than *si* (思, which I translate “thoughts”) which are more often dispassionate. A quotation from Liu Zongzhou (劉宗周, Jishan 戡山, 1578–1645) illustrates this contrast; he 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.” (思而動于欲爲念。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) should not be understood as a concern in the sense described in 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 that” (“I’m concerned that they aren’t here yet”), but fits best with “to be concerned to” (“I’m concerned to get there on time”; “my concern is to ensure all of you get there safely”). Just as I noted earlier that Wang often thinks of *yi* as short-lived episodes (whereas inclinations are typically longer-lasting), so too the reader should bear in mind that Wang often thinks of *nian* as short-lived episodes (whereas concerns are often longer lasting).

²¹In fact I suspect this assumption is true: that Wang held that something is an inclination if and only if it is a motivating concern, and, moreover, that something is a motivating concern if and only if it is a concern. But nothing will turn on the assumption here; everything below could be rewritten to include motivating concerns as a separate class of mental events, in addition to inclinations.

²²There is a clear distinction between the qualities expressed by the terms I translate “right”/“wrong” (是非) and those I translate as “good”/“bad” (善惡), but the exact character of this distinction will not be important for my purposes in this paper. In particular, I will not be assuming that the distinction is or is not the same as the distinction between right/wrong and good/bad as understood by moral philosophers working in English today.

Whenever a motivating concern arises, your mind's *liangzhi* automatically knows it. [If it is good] your mind's *liangzhi* automatically knows that it is good; [if it is bad], your mind's *liangzhi* also automatically knows that it is bad. (QJ 26.1070, Chan (1963, p. 278), Ivanhoe (2009, p. 170))²³

In many further passages, Wang says that a person's *liangzhi* acquires relevant knowledge, no matter how morally corrupt the person has become.²⁴ These passages provide yet more support for the claim that *liangzhi* always knows the qualities of one's mental events. Wang does not hold that only ideally virtuous people experience the exercise of *liangzhi*. He says quite clearly that no matter one's state of virtue or vice, *liangzhi* always knows.²⁵

Wang holds that if a person has a good, right, bad or wrong inclination, their *liangzhi* knows that it is good, right, bad or wrong. Does he also endorse the converse, that if a person's *liangzhi* knows that an inclination is good, right, bad or wrong, then the inclination is their own and it is in fact good, right, bad or wrong? The evidence is less direct, but it suggests he does. First, Wang doesn't explicitly say that if *liangzhi* knows that an inclination is good, right, bad or wrong, it is, but this goes without saying: just as it's not possible to know that green grass is red, it's not possible to *zhi* that a bad inclination is good. Second, Wang holds that a person's *liangzhi* knows that inclinations are good, right, bad or wrong on the basis of a kind of bodily experience of the inclination.²⁶

²³See also the second and third sentences of the "Four Sentence Teaching" (四句教) (IPL 315 QJ 133-4), as well as: IPL 162, QJ 76; IPL 259, QJ 118; IPL 318, QJ 135-6; QJ 8.307. Wang makes similar points in his pre-*liangzhi* period, in the 1515 "Preface to the Old Version of the Great Learning" QJ 7.271. Chen Jiuchuan (陳九川) also makes related points in IPL 201 QJ 102. TODO add 140, connecting those who act without knowing to the mind of right and wrong.

²⁴In "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 7.271). (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 QJ 5.193 (Ching (1972, p. 49)), and emphatically in IPL 152 (QJ 69), IPL 207 (QJ 105), and QJ 27.1112-1113 (Ching (1972, p. 121)). He makes related remarks in: IPL 151 QJ 69; IPL 169 QJ 81-2; IPL 289 QJ 126; QJ 7.298.

IPL 290 (QJ 126) is at first sight in tension with this evidence. There, Wang says that *liangzhi* can know the qualities of these states, not that it does. Moreover, he says that an immediate consequence of *liangzhi*'s awareness of bad emotions is that they will disappear. This passage seems incompatible with the remarks just cited, since if (as Wang says in IPL 207 (QJ 105)) a thief's *liangzhi* knows that they ought not to be a thief, 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.

²⁵In the main text I've only considered inclinations and motivating concerns, but Wang makes related points about *liangzhi*'s knowledge of thoughts (*si* 思) in IPL 169 (QJ 81-2), and emotions or feelings (*qing* 情) in IPL 290 (QJ 126).

²⁶"If you do not rely on your own *liangzhi* to genuinely and practically learn by bodily experience, it is like using a scale without markings to weigh what is light and heavy, or using a mirror that hasn't been opened to reflect what is beautiful and ugly..." (若不就自己良知上真切體認，如以無星之稱而權輕重，未開之鏡而照妍媸 IPL 146, QJ 66). In what follows he connects the use of *liangzhi* in this way directly to a

Since a person cannot have this sort of direct experience of *other people's* inclinations, it is natural to think that this commits him to the claim that a person's *liangzhi* can acquire knowledge in the relevant, direct way only of a their own inclinations.²⁷

In short, Wang holds that a person has a good, right, bad or wrong inclination if and only if *liangzhi* knows that the inclination is good, right, bad or wrong. For our purpose only *liangzhi's* knowledge that inclinations are good will be important in what follows. So the key takeaway from the foregoing discussion, and the first of the two principles which make up my introspective model of genuine knowledge, will be:

Knowing Good A person has a good inclination if and only if the person's *liangzhi* knows that the inclination is good.

In [T3], Wang describes *liangzhi* as acquiring knowledge “automatically” (自). Other passages suggest he holds that *liangzhi* also acquires knowledge effortlessly.²⁸ This idea – that people effortlessly and automatically acquire knowledge that certain mental events are good, right, bad or wrong – might seem unfamiliar on first encounter. But reflecting on features of the conscience as it is commonly understood may help to make it less so. A person's conscience is often credited with automatic and effortless reactions which can be part of acquiring ethically relevant knowledge. For instance, someone who has a spontaneous urge to harm someone else, may simultaneously experience revulsion at what they feel an urge to do. In the right circumstances, this revulsion is naturally described as an exercise of the person's conscience; the person might say “although I wanted to do it, my conscience told me not to” or, if the urge was something they initially felt a need to follow (perhaps it stemmed from righteous anger) they might say “I almost did it, but my conscience told me it was wrong”. These

quotation from Cheng Yi (程頤, Yichuan 伊川, 1033-1107) describing how one is able to distinguish right from wrong (能辨是非).

²⁷Elsewhere, Wang says that *liangzhi* is primarily focused on features of oneself. He writes that it “has nothing to do with others” (是皆無所與於他人者也 QJ 26.1070, Chan (1963, p. 278)) and even allows himself to appropriate a quotation from his usual opponent Zhu Xi (Johnston & Ping (2012, p. 155), Zhu (1983, p. 7)), describing *liangzhi* as “what others do not know, but I know in private” (人雖不知，而已所獨知 IPL 318 QJ 135). Since others cannot acquire this knowledge about me, in particular they cannot acquire it by using their *liangzhi*. And by parallel reasoning it must be that I cannot have this knowledge of *other's* mental lives using *my liangzhi*. So a person can only have this knowledge of their own inclinations.

²⁸In the original passage in which the *Mencius* introduces *liangzhi* (7A.15), the text says that *liangzhi* “does not await reflection before it knows” (不待慮而知) and that *liangneng* “does not await learning before it is able to” (不待學而能). In two passages, Wang affirms that these features belong to *liangzhi* (QJ 26.1070; Chan (1963, p. 278) and QJ 8.311), and, in both of them, goes on to describe *liangzhi* as “the mind which judges right and wrong” or “the mind which approves and disapproves” (是非之心). While the *Mencius* seems to have used these descriptions to indicate that the relevant knowledge and ability were innate, Wang seems to think of them as related to effortlessness, and ties this effortlessness directly to the capacity for judging right and wrong.

deliverances of the conscience are automatic, effortless, uncalled-for reactions, and the testimony of the conscience (its “telling”) is a source of knowledge.²⁹

In contemporary analytic philosophy “introspection” is used as a general label to cover all direct knowledge a person has of their own mind. This technical usage is somewhat broader than the use of the word in ordinary English, where “introspection” is typically reserved for an effortful, conscious process of looking inwards. In this paper, I will employ the technical sense of the word: I will often describe *liangzhi*’s knowledge of the ethical qualities of mental events as “introspective” knowledge, although, as the preceding paragraphs make clear, I believe this knowledge is acquired automatically and effortlessly.

What, if anything, does this discussion of *liangzhi*’s knowledge teach us about genuine knowledge? We can already argue for one important claim: it is possible for a person’s *liangzhi* to know that inclinations are good even when the person does not have genuine knowledge. As we have seen, Wang holds that it is not just the *liangzhi* of sagely people which knows; a person’s *liangzhi* knows that inclinations are good, right, bad or wrong, even when the person is generally morally corrupt, indeed even when they are in the midst of performing vicious bodily actions. But Wang also makes clear in several places that a key component of the doctrine of the unity of knowledge and action is that anyone who is performing vicious bodily actions does not genuinely know filiality, or any other virtue for that matter (*IPL* 5, *QJ* 4; cf. *IPL* 8, *QJ* 7). Together these claims show that Wang is committed to the view that a person’s *liangzhi* can know that an inclination is good even when the person does not have genuine knowledge. This conclusion, then, frames our next question: what is the relationship between the knowledge *liangzhi* always has and genuine knowledge?

²⁹There are many important differences between *liangzhi* and the conscience; here are two. First, Wang holds that *liangzhi* is responsible for seeing and hearing (*IPL* 168 *QJ* 80 (seeing and hearing are all “functions” 用 of *liangzhi*), cf. *QJ* 6.235, Ching (1972, p. 110)) and in some moods even suggests that it is responsible for all knowledge (*QJ* 6.243, Ching (1972, p. 115), though the relevant passage is not translated in her selection; *QJ* 20.871, Ivanhoe (2009, p. 182)); clearly the conscience does not have such broad powers. Second, the conscience appears to know only right and wrong, but Wang clearly holds that *liangzhi* can also know good/bad, and may even hold that it can richer properties such as filiality (see section 5 for more discussion).

Sill, some of the core phenomenology which Wang aimed to capture with his theory of *liangzhi* is similar to what is often associated with the conscience. For some other discussion of the relationship between the conscience and *liangzhi*, see, e.g. 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).

4 The obscuration argument

As a prelude to answering this question directly, I will examine a passage from Wang's important *Questions on the Great Learning*, and in particular a section where Wang addresses the relationship between making inclinations wholehearted (誠意) and "extending knowledge" (致知):

[T4] 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，好之真如好色，發一念而惡也，惡之真如惡惡臭，則意無不誠，而心可正矣。然意之所發，有善有惡，不有以明其善惡之分，亦將真妄錯雜，雖欲誠之，不可得而誠矣。故欲誠其意者，必在於致知焉。°°°

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。是皆無所與於他人者也。故雖小人爲不善，既已無所不至，然其見君子，則必厭然掩其不善，而著其善者，是亦可以見其良知之有不容於自昧者也。

今欲別善惡以誠其意，惟在致其良知之所知焉爾。何則？意念之發，吾心之良知既知其爲善矣，使其不能誠有以好之，而復背而去之，則是以善爲惡，而自昧其知善之良知矣。意念之所發，吾之良知既知其爲不善矣，使其不能誠有以惡之，而覆蹈而爲之，則是以惡爲善，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 sights, and whenever a concern arises and it is hateful [bad], you genuinely hate it as you hate hateful [bad] odors, then all of your inclinations will be wholehearted and your mind can be rectified.³⁰ However, some of the inclinations which arise are good and some are bad. If one did not have a means to understand the distinction between good and bad, and wrongly mixed up true (真) and misguided, then although one wanted to make them [viz. one's inclinations] wholehearted, they cannot successfully become wholehearted. Thus making one's inclinations wholehearted must depend on extending one's knowledge of them...

Whenever a motivating concern arises, your mind's *liangzhi* automatically knows it. [If it is good] your mind's *liangzhi* automatically knows that it is

³⁰"Love lovely sights" and "hate hateful odors" are quotations from the *Great Learning*; see [T6] below. My translation attempts to simulate the fact that "love" is written with the same character (好) as the adjective "lovely" (although they are pronounced differently) and the verb "hate" is written with the same character (惡) as the adjective "hateful" (although they too are pronounced differently). (Above, I have used "bad" for the character I here translate as "hateful".)

good; [if it is bad], your mind's *liangzhi* also automatically knows that it is bad.³¹ It has nothing to do with other people. Thus, although a petty person has become not good, and there is nothing they will stop at, nevertheless when they meet a noble person, they will ashamedly hide the fact that they are not good, and broadcast that they are good. From this one can see that there are some respects in which their *liangzhi* has not allowed itself to be obscured.

Now, if you want to discriminate good and evil in order to make your inclinations wholehearted, this just depends on extending what your *liangzhi* knows about them and nothing more. Why is this? When a [good] motivating concern arises, the *liangzhi* of your mind already knows that it is good. Suppose you do not wholeheartedly love it but instead turn away from it and diminish it. You would then be taking what is good to be bad and obscuring your *liangzhi* which knows that it is good. When a [bad] motivating concern arises, the *liangzhi* of your mind already knows that it is bad. Suppose you do not wholeheartedly hate it but instead backslide and promote it. You would then be taking what is bad to be good and obscuring your *liangzhi* which knows that it is bad. In such cases one says that you know it, but in fact you do not know – how could your inclinations have become wholehearted! [But] now if what *liangzhi* [recognizes as] good or bad is wholeheartedly loved or hated, one's *liangzhi* is not deceived and one's inclinations can be wholehearted. (QJ 26.1070-1, cf. Chan (1963, p. 277-9))

Wang says that making inclinations wholehearted “depends on” (在於) extending knowledge, and also that discriminating good and evil in order to make inclinations wholehearted “depends on” (在) extending knowledge. At the close of the passage he says that if *liangzhi* is not deceived (i.e., presumably, if knowledge is extended), one's inclinations “can be” wholehearted. These remarks might seem most naturally to suggest that having extended knowledge is a necessary condition for having wholehearted inclinations. But Wang often repeats (including at the end of this work) his distinctive view that the four aspects of personal ethical training in the *Great Learning* are really all just different ways of looking at the same task: if one of them is brought to completion, they all must be. He therefore also holds that extending knowledge is sufficient for wholeheartedness of the inclinations. And in fact this claim seems to be in his sights in the itself. The phrase “can be wholehearted” in the last sentence of the excerpt is most naturally read as “the only obstacles to their being wholehearted will be removed” suggesting that Wang means to claim that extended knowledge is sufficient for wholehearted inclinations. And most importantly, as we will see below, his main explicit argument in the third paragraph of the excerpt addresses the sufficiency claim,

³¹This text was also printed above as [T3].

not its converse. These facts suggest that we should here take Wang's "depends on" to mean something stronger than "requires" and that Wang aims to defend the claim that a person has wholehearted inclinations if and only if their knowledge is extended.³²

The argument stated explicitly in the third paragraph of [T4], which I will call the *obscuration argument*, is an argument for an intermediate conclusion on the way to this claim. Here Wang argues that if a person does not wholeheartedly love (好) a good (善) motivating concern, then they do not know that it is good, and similarly that if a person does not wholeheartedly hate (惡) a bad (惡) motivating concern, they do not know that it is bad. The argument runs as follows (in the case of a good motivating concern, which I will focus on throughout):

1. If one does not wholeheartedly love a motivating concern, one turns away from it and diminishes it.
2. If one turns away from and diminishes a motivating concern, one takes it to be bad.
3. If one takes something to be bad, one does not know that it is good.
4. So, if one does not wholeheartedly love a motivating concern, one does not know that it is good.³³

As stated explicitly in the passage, Wang's argument concerns a person's knowledge, and not whether their knowledge is extended. But in context, it is clear that Wang does not mean to deny that the person knows the relevant claim, but only to deny that their knowledge of it is extended. In the second paragraph of [T4] Wang says that *liangzhi* always knows. In the passage as a whole, Wang aims to elucidate precisely the contrast between the knowledge *liangzhi* always has and the knowledge it would have if it were extended. This contrast concerns forms of *liangzhi*'s knowledge,

³²Since Wang does not discuss the unity of knowledge and action explicitly in this passage, it does not follow from my principle for demarcating cases of "extended knowledge" which describe genuine knowledge that his uses of "extended knowledge" in this passage describe genuine knowledge, and accordingly in what follows I will not assume that they do. See below n. 39 for more discussion.

³³It may be helpful to think of the second and third premises of Wang's argument by comparing them to the premises of 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:

- 2 If one wants that p , one believes that it's good that p .
- 3 If one believes that p , one does not know that $\neg p$.
- 4 If one wants that p one does not know that it's not good that p .

but Wang suggests a related contrast is reproduced among forms of the *person's* knowledge. Again in the second paragraph of the excerpt, Wang says that petty or vicious people recognize their faults and hide them in front of virtuous people. Wang takes this to show that even vicious people know that they are not good in virtue of the exercise of their *liangzhi*. But of course Wang would not credit vicious people with *extended* knowledge; extended knowledge is something only virtuous people have. Moreover, at the end of the excerpt, Wang admits that in a case where a person fails to wholeheartedly love their good inclinations, we do *say* that the person knows. He here seems to recognize a divergence between his own way of speaking of (extended) knowledge, and the form of knowledge described by ordinary uses of the word "know". Elsewhere, too, Wang implicitly concedes that someone may be said to know even though they do not have genuine or extended knowledge (*IPL* 5, *QJ* 4; *IPL* 138, *QJ* 55). A person who does not wholeheartedly love their good inclinations does know in some sense that those inclinations are good, they just do not (really) know that they are. In sum, Wang's argument should be understood as targeting whether the person in question has extended knowledge, not whether they have knowledge at all. The third premise of the argument should be understood as "If one takes something to be bad, one does not have extended knowledge that it is good"; and the conclusion as "If one wholeheartedly loves a good motivating concern, one does not have extended knowledge that it is good."

Thus, the third premise of the argument articulates a necessary condition for knowledge that an inclination is good to be extended, namely, that the person not take the inclination to be bad. In what sense might "taking what is good to be bad" (以善爲惡) prevent a person from having an elevated form of knowledge that the inclination is good?

Consider a native Chinese speaker who pronounces the Pinyin sound "zh" in the standard way, by putting their tongue against the back of their palette, but who, when they reflect on their practice in pronouncing this sound, comes to believe that the sound is produced by putting their tongue at the front of their mouth against their lower teeth. If there could be any people who both know a claim, and believe the negation of that claim, this person is a good candidate: they know – as evinced by their ability to reliably, intentionally produce the sound correctly – that the sound is produced by putting one's tongue at the back of one's palette. But they also believe – as a result of reflecting on their practice – that the sound is not produced in this way. Now compare this person ("the conflicted person") to someone ("the unconflicted person") who can produce the sound but who also, on reflection, comes to believe truly that the sound is produced

by putting their tongue at the back of their palette. The unconflicted person's state of mind can be said to be better than the conflicted person's state of mind in at least one distinctively epistemic or doxastic respect, since the conflicted person suffers from a form of doxastic conflict which the unconflicted person does not suffer from. In a grandiose mood, one might say that, in virtue of this fact, the unconflicted person has an elevated form of knowledge by comparison with the conflicted person.

Wang's idea in our passage seems to be related. He considers a structurally parallel contrast between people in two different conditions. In one condition, a person knows that an inclination is good and wholeheartedly loves it; in another, a person knows that an inclination is good, but does not wholeheartedly love it – and, as a result, takes it to be bad. A person who knows that an inclination is good, but also “takes what is good to be bad” suffers from something like the doxastic conflict described above. Presumably, the person who wholeheartedly loves their good inclination does not suffer from such conflict.³⁴ Wang articulates the contrast between these two conditions in terms of the person's knowledge; his remarks suggest that if a person suffers from this conflict they have a lower grade of knowledge i.e. knowledge which is not extended. But the idea that there is an important contrast between the two different conditions does not depend on this way of articulating the thought. Without speaking of grades of knowledge, we could instead paraphrase Wang's idea by saying that the state of mind of a person who wholeheartedly loves a good motivating concern is better in a distinctively epistemic or doxastic respect than a person who does not, because the person whose love is not wholehearted suffers from a form of doxastic conflict which the person whose love is wholehearted does not suffer from.

The first and second premises of the argument require much less discussion. The first states that if one does not wholeheartedly love a motivating concern, one turns away from it and diminishes it.³⁵ Wang's point seems to be that, since the response

³⁴The fact that Wang connects “taking what is good to be bad” and “taking what is bad to be good” to the “self-deception” described by the *Great Learning* (see below [T6]) may provide further support for the idea that Wang means to emphasize doxastic conflict in his discussion (cf. also IPL 138, QJ 55; IPL 171, QJ 84). Just as a person who suffers from self-deception about the extent of their own accomplishments may be said to know, deep down, that they have not achieved much, in spite of the fact that they outwardly seem to believe they have accomplished a great deal, so too the person described by Wang's gloss on “self-deception” in the *Great Learning*, knows the goodness of a given inclination (via *liangzhi*) but also takes it to be bad.

³⁵In the passage Wang does not make this claim directly; he says that “if one does not wholeheartedly love a motivating concern, but instead turns away from it and diminishes it”. On its own this sentence could be read as saying not that wholeheartedly loving an inclination on the one hand and turning away and diminishing it on the other are exhaustive alternatives but only that turning away and diminishing an inclination is a special, extreme form of failing to wholeheartedly love it. This interpretation would seem on an even stronger footing if one reads (as is linguistically possible) the words I have translated as

good inclinations warrant is wholehearted love, anything short of that response amounts to diminishing their status, and turning away from them.³⁶

The second premise connects an affective feature of a person's response to an inclination ("turning away") with the more doxastic "taking". Wang says that if one turns away from and diminishes an inclination, one takes it to be bad. He says nothing in this passage to defend this thesis. He may have seen it as simply obvious that there is a doxastic component to these affective reactions. For instance, he might have seen it as obvious that loving something requires taking it to be good, and correspondingly that turning away from and diminishing must also have a doxastic component.³⁷

These premises entail the conclusion of the obscuration argument, which says that

"diminish" (去) and "promote" (爲) as "eliminate" and "enact" respectively. But in the context of Wang's argument here, this reading and the associated translation are disfavored. Wang clearly takes his argument to show that *liangzhi's* powers are not fully exercised if one does not wholeheartedly love a good motivating concern. If he thought that turning away and diminishing a good motivating concern was just a special case of failing to wholeheartedly love it, there would be a gaping hole in his argument: he would have failed to show that *liangzhi's* knowledge would also be degraded if the person exhibited different, less extreme ways of failing to wholeheartedly love the relevant inclination.

³⁶In a moment I'll argue that Wang holds that good inclinations are wholehearted if and only if they are wholeheartedly loved. In section 2, I suggested that inclinations are wholehearted if and only if there are no inclinations which conflict them. Given these two claims, if a person's good inclination is not wholeheartedly loved, it is not wholehearted, and hence the person has a conflicting inclination. On this picture, Wang might be saying, in particular, the tug in a different direction from the conflicting inclination amounts to "turning away from and diminishing" the good inclination.

³⁷It is possible that this connection between turning away and taking stemmed from a deeper commitment on Wang's part about the character of *liangzhi's* response to inclinations:

[T5] 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)

In this passage, the characters I have elsewhere translated as "right" and "wrong" (*shi fei* 是非) (which are the same as the nouns "rightness" and "wrongness") are used as transitive verbs. I have rendered them "judge right" and "judge wrong" to mark the fact that it is the same character, but they might also be translated as "approve" and "disapprove". The passage clearly says that *liangzhi's* approving an inclination or judging it to be right *just is* loving that inclination. It leaves open a variety of positions about how this "approving" or "judging to be right" is related to what Wang elsewhere describes as *liangzhi's knowing*. An intriguing hypothesis is that Wang takes *liangzhi's* knowledge of the rightness of an inclination to be identical to *liangzhi's* approving the inclination. This idea can seem natural when one reflects on some phenomenology associated with the conscience: although we can speak of the conscience "telling" us that something is wrong or right, and so conferring knowledge about it, such a "telling" may feel internally more like an affective reaction. Wang could have been led by such considerations to identify *liangzhi's* recognition of the ethical qualities of mental events with its loving or hating them. And if he did identify the knowing and apt loving of *liangzhi* (on whatever grounds), it would be natural for him also to take a similar position about the relationship between the mistaken "taking" and inapt "turning away" he describes in [T4]. If the taking and the turning away are the very same thing in this case, there would be no mystery as to why the presence of the one would imply the presence of the other.

if a person does not wholeheartedly love an inclination, then they do not have extended knowledge that it is good. By contraposition, this conclusion is equivalent to the claim that if a person has extended knowledge that an inclination is good, then they wholeheartedly love it. This claim almost establishes a connection between extended knowledge and wholehearted inclinations. And the missing step is not hard to see. In the passage, Wang repeatedly alludes to one of his favorite passages from the *Great Learning*:

[T6] 所謂誠其意者，毋自欺也，如惡惡臭，如好好色

What is called making the inclinations wholehearted is not deceiving oneself.

It is like hating a hateful [bad] odor, like loving a lovely sight.

He transposes the *Great Learning's* talk of hating hateful [bad] odors, and loving lovely sights to a person's metacognitive reactions to their inclinations; he speaks of hating hateful [bad] inclinations and loving good inclinations. He implicitly relies on the claim – which he seems to take to be supported by [T6] – that all of a person's inclinations are wholehearted if and only if they wholeheartedly love their good inclinations and wholeheartedly hate their bad ones. Given Wang's commitment to this claim about all inclinations, it is plausible that he also holds a principle I will call "Wholehearted Love": that an inclination to perform a good action (i.e. a good inclination) is wholehearted if and only if it is wholeheartedly loved. Given this equivalence, we can move from the conclusion of the obscuration argument to the claim that if a person has extended knowledge that an inclination is good, then they have a wholehearted inclination to perform a good action.

I noted at the outset that our goal was to understand Wang's arguments for the claim that a person has extended knowledge if and only if their inclinations are wholehearted. The obscuration argument (together with Wholehearted Love) most obviously targets the claim that if knowledge is extended, inclinations are wholehearted. But as I noted at the outset, Wang's remarks about his goal in the passage are most readily understood as focused on the converse of this claim, i.e. that if inclinations are wholehearted, then knowledge is extended. While this disconnect may seem puzzling at first sight, it is easy to see how the argument Wang does give can be strengthened into an argument that he would have accepted, and which would support the claim he says he has in view. For, it is natural to think that Wang would accept the converses of each of the three premises of the obscuration argument, and thus accept the converse of its conclusion as well. For the first premise: Wang says explicitly that if one does not wholeheartedly love an inclination, one turns away from it and diminishes it. But he

clearly also thinks that if one turns away from and diminishes an inclination, one does not wholeheartedly love it. Similarly, for the second premise: Wang says that if one turns away from and diminishes an inclination, one takes it to be bad. But it is natural to think that he would also hold that taking a good inclination to be bad is a way of turning away from it and diminishing it, and thus that if one takes a good inclination to be bad, one turns away from it and diminishes it. Finally, for the third premise: Wang says that if one takes good to be bad, then *liangzhi's* knowledge that the inclination is good is obscured. But it is natural to think that he would hold that states which entail this kind of conflicting taking are the only way that *liangzhi's* knowledge that the inclination is good would be obscured, and thus that if *liangzhi's* knowledge that the inclination is good is obscured, then one takes what is good to be bad. If these remarks are correct, and Wang was also committed in this way to the converses of each of the premises of the obscuration argument, then he would also have been committed to the converse of its conclusion, that is, to the claim that if a person wholeheartedly loves an inclination to perform a good action, they have extended knowledge that it is good. Given the principle introduced in the previous paragraph (that a good inclination is wholehearted if and only if it is wholeheartedly loved), this entails that if a person has a wholehearted inclination to perform a good action, then they have extended knowledge that it is good.

In this section, we have seen arguments for the claim that a person has extended knowledge that an inclination is good if and only if they have a wholehearted inclination to perform a good action. In the next section I will provide an account of genuine knowledge which will help us to move from this claim to Genuine/Wholehearted, the claim that a person genuinely knows filiality if and only if they have a wholehearted inclination to perform a filial action.

5 The introspective model

To introduce my account of genuine knowledge, it is worth considering what an account of genuine knowledge should aim to achieve. In particular, it is worth considering why Genuine/Wholehearted on its own does not provide an adequate account of genuine knowledge. An important reason is this: Genuine/Wholehearted describes only the conditions under which a person has genuine knowledge; it does not say what genuine knowledge is. The principle thus leaves open the possibility, for example, that 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 a triviality

masked as a substantive truth by a misleading definition. 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 knowledge (as that term is ordinarily understood), never mind an elevated form of knowledge. An account of genuine knowledge must not just say when people have genuine knowledge, but also what genuine knowledge is.

To clarify the difference between these two questions, it may be useful to consider first a simple, incorrect interpretation. 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 moment, let's assume that they are. Consider:

Simple Knowledge Something is an episode of a person's genuinely knowing filiality if and only if it is an inclination of theirs to perform a filial action, which is wholeheartedly loved.

Given natural background assumptions, Simple Knowledge and Wholehearted Love (see text following [T6]) entail Genuine/Wholehearted.³⁸ 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 a filial action.

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 [T2], 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

³⁸The assumptions are: (i) that a person genuinely knows filiality if and only if something is an episode of their genuinely knowing filiality; (ii) inclinations to perform filial actions which are wholeheartedly loved are inclinations to perform good actions; (iii) wholehearted inclinations to perform filial actions are inclinations to perform good actions. Recall that Wholehearted Love states that: an inclination to perform good action is wholehearted if and only if it is wholeheartedly loved.

Suppose a person genuinely knows filiality. By (i) something is an episode of their genuinely knowing filiality. By Simple Knowledge, this is an inclination to perform a filial action which is wholeheartedly loved. By (ii) this inclination is a good inclination. Since it is wholeheartedly loved, by Wholehearted Love it is wholehearted. So, the person has a wholehearted inclination to perform a filial action.

Suppose a person has a wholehearted inclination to perform a filial action. By (iii) the inclination is good; by Wholehearted Love it is wholeheartedly loved. By Simple Knowledge, this inclination to perform a filial action which is wholeheartedly loved is an episode of the person's genuinely knowing filiality. By (i), the person genuinely knows filiality.

that what it is to know that a right motivating concern is right, or to know that a good motivating concern is good just is to have that motivating concern. But he doesn't. In [T4], Wang 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 motivating concern would obscure *liangzhi* which knows that that the motivating concern is good. His point is that this distinct mental event of knowing that the concern is good could not in the relevant circumstances lead to or constitute extended knowledge. As noted in section 2 (see references in n. 14), Wang repeatedly ties extending *liangzhi* to the unity of knowledge and action. These passages make it natural to suppose that episodes of genuine knowledge of filiality at least have episodes of *liangzhi*'s knowing that the relevant concern is good (or right) *as parts*. Given this supposition, Wang's clear indications in [T4] that *liangzhi*'s knowledge is distinct from a first-order inclination, would imply that the simple model is incorrect.

This last argument against the simple model suggests an alternative hypothesis: that Wang identifies genuine knowledge of filiality not with an inclination to perform a filial action, but with *liangzhi*'s introspective knowledge that such an inclination is good. Wang describes in many places how *liangzhi* must be extended to achieve an elevated state of knowledge connected to the unity of knowledge and action (see again n. 14 for references). As we saw in section 3, when Wang describes *liangzhi*'s knowledge, he often speaks of it as knowing that particular mental events are good, right, bad or wrong. In one passage, he directly connects his conception of *liangzhi* as "the mind which judges right and wrong" to the unity of knowledge and action (QJ 5.211 Ching (1972, p. 68-9)). Together, these pieces of evidence suggest the following thesis:

Introspective Knowledge Something is an episode of a person's genuinely knowing filiality if and only if it is an episode of their *liangzhi*'s knowing that an inclination – which is in fact an inclination to perform a filial action, and is wholeheartedly loved – is good.³⁹

³⁹I am uncertain whether Wang thinks there is genuine knowledge of bad ethical qualities. It is clear that Wang thinks *liangzhi* knows that inclinations are wrong or bad just as much as it knows that they are good or right. But allowing genuine knowledge of badness would allow that the variant of Genuine/Wholehearted could fail in either direction if "filiality" were replaced with "wrongness": the conditions under which a person would have genuine knowledge of a bad quality are naturally taken to be conditions under which the person in fact was removing a bad inclination; it would certainly not be a case of having a wholehearted inclination to be perform a bad action. Since failures of Genuine/Wholehearted conflict with the general picture Wang seems to be developing, that is a point against taking there to be genuine knowledge of bad ethical qualities. If one holds that the "extended knowledge" in [T4] is indeed genuine knowledge, then Wang would be committed to there being genuine knowledge of bad ethical qualities;

Introspective Knowledge and Knowing Good make up the introspective model of genuine knowledge. Under natural assumptions, the theses, together with Wholehearted Love, entail Genuine/Wholehearted.⁴⁰ 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 that the person's relevant inclination is good.

An example may help to illustrate the contrast between the introspective model and Simple 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. Let us suppose that the inclination is good, and so (given Knowing Good) this person's *liangzhi* knows that it is good. Moreover, let us suppose that the person wholeheartedly loves the inclination.

Given the relevant supporting assumptions (see nn. 38 and 40), both Simple Knowledge and the introspective model imply Genuine/Wholehearted; they agree that the person described here genuinely knows filiality. But they differ on which mental events of the person are episodes of genuine knowledge. According to Simple Knowledge (which we saw to conflict with the texts) it would be the inclination to perform a filial action itself. According to Introspective Knowledge, by contrast, the event of genuinely knowing filiality is the event of *liangzhi's* knowing that the filial inclination is good. This event is a distinct one from the filial inclination itself.⁴¹

that is also an interesting view, and worth developing in more detail, though I won't consider it further here.

⁴⁰The assumptions are (i) and (iii) from n. 38, together with (iv) a person's *liangzhi* knows that an inclination is good if and only if something is an episode of their *liangzhi's* knowing that it is good.

Suppose a person genuinely knows filiality. Then by (i) something is an episode of their genuinely knowing filiality. By Introspective Knowledge, the episode of their genuinely knowing filiality is an episode of their *liangzhi's* knowing that an inclination – which is in fact an inclination to perform a filial action and wholeheartedly loved – is good. By (iv) their *liangzhi* knows that the inclination is good. By Knowing Good, the inclination is good. By Wholehearted Love, it is wholehearted. So the person has a wholehearted inclination to perform a filial action.

Suppose a person has a wholehearted inclination to perform a filial action. By (iii) the inclination is good. By Wholehearted Love it is wholeheartedly loved. By Knowing Good, the person's *liangzhi* knows that it is good. By (iv) something is an episode of the person's *liangzhi* knowing that an inclination (which is in fact an inclination to perform a filial action, and is wholeheartedly loved) is good. By Introspective Knowledge, this is an episode of their genuinely knowing filiality. By (i) they genuinely know filiality.

⁴¹There is a third thesis, Total Knowledge, which I consider at length in the companion paper, [Lederman \(2020a\)](#). 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 rules out this claim: even if the person in our example is perceiving their

It is time now to see how the introspective model answers the challenge introduced at the start of the paper. The challenge can be seen as having two parts: first, to show that genuine knowledge of filiality is recognizably a form of knowledge – and in particular, as argued in section 2, objectual knowledge of filiality; second, to elucidate the sense in which genuine knowledge is an elevated form of knowledge of filiality, and not simply stipulated to be “whatever knowledge happens to be unified with action”. I will take these points in turn.

The introspective model clearly answers the first part of the challenge as it was formulated in the introduction. According to Introspective Knowledge, genuine knowledge is a form of knowledge: it is *liangzhi*'s knowing that an inclination is good. But it is less obvious how it answers the more refined version of this challenge presented at the end of section 2: of explaining why this form of knowledge is reasonably thought of as objectual knowledge of filiality. As a prelude to an answer, a few remarks about the quality of filiality are in order. Wang fairly clearly thinks that actions are filial, if they are, because the person has a “mind which is filial to their parents” (孝親之心).⁴² Since the filiality of the mind explains the filiality of actions, there is a sense in which the mind is the primary bearer of the quality of filiality. Wang does not describe inclinations themselves as filial, but elsewhere he identifies “the mind which...” with an inclination, and it is natural to think he takes the mind which is filial to be a filial inclination.⁴³ Accordingly, I will assume that what I have been calling “inclinations to perform filial actions” to this point just are filial inclinations.

Given Wang's idea that filiality is primarily instantiated in the mind, it is natural to think that he would hold that the most direct cognitive relation one could have to the property of filiality would be a form of introspective knowledge. What would this introspective knowledge be? The most obvious answer is: knowing that inclinations are filial. But Wang seems to hold that *liangzhi* does not know that inclinations are filial (or

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 that an inclination is good. See below for further discussion.

⁴²IPL 133, QJ 48, where he makes the same point about the relationship between actions which are conscientious toward one's ruler and “the mind which is conscientious toward the ruler” (忠君之心). Cf. IPL 3, QJ 2-3 and IPL 135, QJ 50-1.

⁴³In IPL 132 QJ 7, Wang says that “the mind which desires food is an inclination” (欲食之心即是意), and that “the mind which desires to travel is an inclination” (欲行之心即是意).

Why then would he not describe them as filial? Here is one hypothesis. Wang holds that it is misleading to describe such inclinations as filial when they are not wholehearted, since that description might suggest that people with filial inclinations already count as filial, whereas only a person with wholehearted inclinations counts as filial. However, if a person does have a wholehearted inclination to perform a filial action, they count as having “a mind which is filial”, and this is a better description since it is less likely to lead to confusion.

respectful, conscientious, humane or compassionate); it only knows that inclinations are good or bad, right or wrong.⁴⁴ But it is not as though Wang holds there is another, distinct faculty, which has more direct introspective knowledge of filiality: *liangzhi*'s knowledge that an inclination (which is in fact filial) is good is plausibly the most direct sensitivity a person can have to the filiality of the inclination. This most direct sensitivity is then reasonably called "knowledge of filiality". This idea is worth expanding on for a moment. On this picture, *liangzhi*'s knowledge that an inclination (which is in fact respectful) is good would be knowledge of respect; its knowledge that an inclination (which is in fact conscientious) would be knowledge of conscientiousness, and so on. The difference between knowing filiality and knowing respect would thus not be in the content of *liangzhi*'s knowledge – *liangzhi* only knows that the inclination in question is good – but rather in whether the relevant inclination is in fact filial or respectful. An example may help with this point: suppose that on two different days, a person sees two identical hats bobbing along in the same way in a crowd. If the hats belong to two different people, then in spite of the fact that what the person saw may have been in some sense exactly the same, they count as having seen two different people. Similarly, although what a person's *liangzhi* knows is exactly the same in the case of a filial or respectful inclination (namely, that the inclination is good), the person counts as knowing filiality by virtue of knowing that an inclination (which is in fact filial) is good, and counts as knowing respect by virtue of knowing that an inclination (which is in fact respectful) is good. Introspective Knowledge fits into this general picture. It says, in essence, that if the inclination is not only filial, but also wholehearted, then *liangzhi*'s knowledge that it is good will not just be knowledge of filiality, but genuine knowledge of filiality.

This brings us to the second part of the challenge: in what sense is the knowledge of filiality an elevated form of knowledge, if the inclination in question is wholehearted? In the previous section we saw that Wang draws an important distinction between, on the one hand, a person whose *liangzhi* knows that an inclination is good and who wholeheartedly loves that inclination, and, on the other, a person whose *liangzhi* knows that an inclination is good, but does not wholeheartedly love the inclination on the other. Wang claims that since a person who does not wholeheartedly love their good inclination suffers from a certain kind of motivational conflict (they "turn away from and diminish" their good concern), they also suffer from a form of doxastic conflict, "taking

⁴⁴In *IPL 8 (QJ 7)*, Wang does say that the mind (and in context quite clearly *liangzhi*) "knows filiality", "knows respect" and "knows compassion". But it is not clear in that passage whether he has in mind knowledge of qualities of mental events, or something more general.

what is good to be bad". Wang proposes that a person who suffers from such doxastic conflict cannot have extended knowledge that the relevant inclination is good; the doxastic conflict degrades their knowledge.

This paradigm applies to gradations in a person's propositional knowledge that an inclination is good. But Introspective Knowledge allows us to extend the paradigm to objectual knowledge of filiality. I suggested a moment ago that in general knowledge of filiality (whether or not it is genuine) should be identified with *liangzhi's* knowledge that an inclination (which is in fact filial) is good: this knowledge is the most direct sensitivity one can have to the property of filiality as it is instantiated in the mind. But given that the episode of knowing filiality is identical to the episode of knowing that the inclination is good, if the knowledge that the inclination is good is degraded by doxastic conflict, it is natural to think that the knowledge of filiality will be degraded too. If you saw a person by virtue of seeing their hat, but you didn't see their hat very clearly, then you didn't see them very clearly. Similarly, if a person knows filiality by knowing that an inclination (which is in fact filial) is good, but their knowledge that the inclination is good is not of the highest form, then their knowledge of filiality is also not of the highest form.

In closing this section, I want to consider and respond to one objection to my interpretation. In perhaps the most famous discussion of the unity of knowledge and action (IPL 5, QJ 4), Wang is asked to discuss an example related to examples of *akrasia*. Wang says that the people in this example – who know that they ought to be filial but fail to act filially – do not have genuine knowledge. He makes this point by quoting [T6], saying: "The *Great Learning* points to genuine knowledge and action for people to see. It says they are 'like loving lovely sights and hating hateful odors'" (故《大學》指個真知行與人看，說『如好好色，如惡惡臭』。IPL, 5, QJ, 4).⁴⁵ This passage has led many interpreters to take perceiving features of the environment to be important to genuine knowledge. Indeed, the position that perception of the environment can at least sometimes be a component of genuine knowledge has a claim to be orthodoxy among English-language interpreters of Wang Yangming.⁴⁶

⁴⁵It is worth noting is that, in this classic discussion of genuine knowledge, Wang appeals to the *Great Learning's* discussion of wholehearted inclinations to illustrate what genuine knowledge is like. This fact provides strong further support for the focus in this paper on Genuine/Wholehearted.

⁴⁶See, for remarks along these lines, Nivison (1973, 132) (reprinted in Nivison (1996, p. 243)), Nivison (1973, p. 134) (reprinted in Nivison (1996, p. 244)), (Ivanhoe, 2002, p. 99), Ivanhoe (2009, p. 113), Ivanhoe (2011, p. 274), Angle (2005), Angle (2010). Cua (1982, p. 9-14) does not take these examples to be examples of genuine knowledge, but does hold that perception of the environment is a component of genuine knowledge. The position of Shun (2011) is subtler. I discuss the the views of all these authors in much

The introspective model is incompatible with this orthodox position. A person's perceiving features of their environment is a different event from their *liangzhi's* knowledge of the ethical qualities of their inclinations. It is also not *part* of the event of their *liangzhi's* knowing the quality of their inclinations.

Should we give up the orthodoxy or the introspective model? Let us examine more closely the evidence on which the orthodoxy is based. In our passage, Wang does not say unambiguously that the examples from the *Great Learning* are instances of genuine knowledge. The examples are introduced by the word *ru* 如, which *can* mean "for example", but can also (equally naturally) mean "like", introducing an analogy or simile. Those who use this passage to argue that perception of the environment can be part of genuine knowledge understand the word *ru* as "for example" here. But there is strong independent evidence against this interpretation. In the following quotation, Wang responds to a correspondent who has argued that the passage from the *Great Learning* describes how one should respond when confronted with a beautiful sight:

[T7] 人於尋常好惡，或亦有不真切處，惟是好好色，惡惡臭，則皆是發於真心。……大學是就人人好惡真切易見處，指示人以好善惡惡之誠當如是耳，亦只是形容一「誠」字。今若又於好色字上生如許意見，却未免有執指為月之病。

Some of peoples's ordinary loves and hates are not genuine and practical (真切). But loving lovely sights, and hating hateful [bad] odors all arise from the genuine mind...The *Great Learning* relies on the fact that it is easy to recognize the genuineness and practicality of love and hate that everyone has, to teach people what the wholeheartedness with which one loves goodness and hates badness [hatefulness] should be like. It just describes the one character "wholehearted". [But] now by the way in which you again put forth so many opinions on the words "beautiful sights", it seems you have not avoided the mistake of "taking the finger to be the moon". (QJ 5.218, cf. Ching (1972, p. 91))

"Take the finger to be the moon" is a saying which describes someone who confuses the pointer with the object pointed at. According to Wang, his correspondent confuses the analogy (loving lovely sights) with what the analogy aims to illustrate (the wholeheartedness with which one ought to love good), thereby taking the finger to be the moon.⁴⁷ Loving lovely sights, says Wang, is not an example of wholehearted

more detail in "Perception, Know How and Genuine Knowledge".

⁴⁷In a number of other passages, Wang also ties the passage from the *Great Learning* to a person's wholehearted love of the good (IPL 229, QJ 110-11, IPL 318 QJ 135-6, and also in [T4] above; in QJ 7.276-8 Wang speaks of *liangzhi* as loving filiality and other virtues. This was also Zhu Xi's reading of the passage; see e.g. Johnston & Ping (2012, p. 155), Zhu (1983, p. 7).

inclinations; it is simply an example which illustrates what wholeheartedness is like. It is natural to think that Wang would have said the same thing about the examples as illustrative of genuine knowledge: the examples illustrate the connection between knowledge and action; they are not themselves examples of genuine knowledge. Interpretations which ascribe to Wang the view that perception is a component of genuine knowledge by taking this “*ru*” 如 to mean “for example”, ignore Wang’s own advice. They seem to mistake the finger Wang used, for the moon.⁴⁸

6 Conclusion

In section 2, I set the stage for the introspective model, arguing that an account of genuine knowledge would make a major step toward understand the unity of knowledge and action if it could be shown to vindicate a connection between genuine knowledge and wholehearted inclinations. In particular I suggested we aim to vindicate:

Genuine/Wholehearted A person genuinely knows filiality if and only if they have a wholehearted inclination to perform a filial action.

In section 3, I argued that Wang accepts:

Knowing Good A person has a good inclination if and only if the person’s *liangzhi* knows that the inclination is good.

But, I argued, *liangzhi*’s automatic knowledge that good inclinations are good, does not invariably amount to genuine knowledge. In section 4, I developed this point by considering the obscuration argument, where Wang explicates a contrast between different forms of knowledge. I argued that he does characterizes this contrast in terms of the absence of a particular form of doxastic conflict, which he ties, in turn, to whether

⁴⁸Can there be non-ethical objects of genuine knowledge on the present view? The interpretation here does not force a stance on this much-discussed question (see e.g. [Frisina \(1989\)](#) and [Cua \(1982\)](#)). The general picture described by the introspective model is incompatible with the opening examples from *IPL* 5 (*QJ* 4) (along with the examples of knowing one’s soup, knowing one’s clothes and knowing the road one will travel on from *IPL* 132 (*QJ* 46-7)) being examples of, rather than analogues for, genuine knowledge. But that does not mean it forces there to be no non-ethical examples at all. For instance, later in *IPL* 5, Wang shifts his examples from colors and sights to hunger, cold and pain. Those examples are all plausibly states which are perceived by an “inner sense”. The letter of Introspective Knowledge, which describes knowledge that an inclination is good, does not apply to these cases, but the spirit of the introspective model is compatible with taking them to be examples of genuine knowledge, and perhaps they should be. There are also tantalizing discussions also of knowing pains and itches, which could also be thought of as genuine knowledge, but again may also simply be analogues to illustrate Wang’s basic points (*IPL* 144 *QJ* 65; “Response to someone who asked about *liangzhi* (Two Verses)” *QJ* 20.871, [Ivanhoe \(2009, p. 182\)](#)).

a person's inclinations are wholehearted. Building on this discussion, in section 5 I proposed that Wang furthermore holds:

Introspective Knowledge Something is an episode of a person's genuinely knowing filiality if and only if it is an episode of their *liangzhi's* knowing that an inclination – which is in fact an inclination to perform a filial action, and is wholeheartedly loved – is good.

I showed how Knowing Good and Introspective Knowledge, together with some natural background assumptions (see n. 40), entail Genuine/Wholehearted. Moreover, I described how they answer the challenge stated in the introduction: to explain how genuine knowledge can be thought of as an elevated form of knowledge, and, in particular, knowledge of filiality. *Liangzhi's* knowledge that an inclination (which is in fact filial) is good is the most direct cognitive sensitivity a person can have to the property of filiality as it is instantiated in their mind. This knowledge is therefore reasonably called knowledge of filiality. Such knowledge has the elevated status of being genuine when the propositional knowledge it is identical to is not degraded by the presence of the doxastic conflict mentioned above. A person who knows filiality but does not genuinely know it, takes their filial, good inclination to be bad. This conflicting “taking” degrades their knowledge that it is good, and thereby also their knowledge of filiality.

One might caricature the introspective model by saying that, according to it, genuine knowledge is an inefficacious and intrinsically valueless prize superadded to the victory of virtuous action. This 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 one's own mental events (which presumably help to guarantee that one is acting virtuously). 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 (*IPL* 5, *QJ* 5; *IPL* 133, *QJ* 47-8; *IPL* 136, *QJ* 52; *QJ* 6.233). To describe the prize as valueless or inefficacious would thus in an important sense be to describe the victory in the very same terms.

This caricature does, however, help to illustrate how starkly Wang differed from his Song dynasty predecessors as he understood them. According to Wang, Cheng Yi (程頤, Yichuan 伊川, 1033-1107) and Zhu Xi had held that ethical knowledge facilitates

virtuous action in part through its role in deliberation. On this Cheng-Zhu orthodoxy as Wang interpreted it, knowledge came first and virtuous action later (*IPL* 133, *QJ* 48). The caricature helps us to see how Wang replaced this knowledge-first position with one on which knowledge, in a certain sense, comes last.

Say that a person is *akratic* if and only if they know that an action is not among the best available to them but they voluntarily do it nevertheless. Wang's doctrine of the unity of knowledge and action is sometimes said to amount to the denial of the possibility of *akrasia*. In closing I want to examine this question in light of the introspective model I have developed.

To examine this question, let us suppose that Wang holds a person has a wholehearted inclination to perform a filial action if and only if they are acting filially (see above n. 5). If Wang endorses this claim, then given his commitment to Genuine/Wholehearted, he is committed to the claim that a person genuinely knows filiality if and only if they are acting filially, and similarly for virtues other than filiality. So he holds that if a person has genuine knowledge of a good ethical quality, they are acting in a way that exhibits that quality: it is not possible to act badly, and at the same have genuine knowledge of a good ethical quality. This claim 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 not among the best available to a person). He also emphasizes genuine knowledge of good qualities, whereas whether someone is *akratic* depends on their knowing that an action is *not* among the best actions available to them. Finally, Wang focuses on genuine knowledge, whereas whether a person is *akratic* depends on what they just plain know.

In fact, it is natural to think that Wang would have held that *akrasia* is absolutely pervasive, and that any person who voluntarily performs a bad action suffers from *akrasia*. Whenever a person voluntarily acts badly, it is natural to think they act on the basis of a bad inclination. Whenever a person has a bad inclination, Wang holds that their *liangzhi* knows that it is bad. So, Wang seems committed to the claim that if a person voluntarily acts badly or wrongly, their *liangzhi* knows that their inclination (and presumably the action they are performing) is wrong or bad. Wang does not speak in terms of "best available actions", but he also does not consider cases like ethical dilemmas where bad actions are sometimes claimed to be the best actions available to a person. Perhaps he would have held that an action is not among the best available to a person if and only if it is bad. If he did, it would then be natural for him to think that a person who knows that an action is bad, knows that the action is not among the best

available to them. So he would be committed to the claim that anyone who voluntarily performs a bad action suffers from *akrasia*.

These remarks help to bring out an important theme. Wang holds that the conscience-like faculty of *liangzhi* always knows that inclinations are good, bad, right or wrong if they are. The ever-present knowledge of *liangzhi* allows him to reduce substantive evaluation of people's mental states to the evaluation of their coherence. Whether or not one accepts my more conjectural remarks in the previous paragraph, it is clear that Wang was close to a position on which any voluntary bad action counts as *akrasia* – the substantive badness is a form of incoherence. And, as we have seen at length, on the introspective model, whether a person acts virtuously (or at least, whether their inclinations are wholeheartedly virtuous) can be characterized in terms of whether they do or do not exhibit a form of doxastic incoherence.

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.
- 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.
- 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.
- Chu, Hung-lam. 1988. The Debate Over Recognition of Wang Yang-ming. *Harvard Journal of Asiatic Studies*, 48(1), 47–70.
- 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.
- Geaney, Jane. 2002. *On the epistemology of the senses in early Chinese thought*. Vol. 19. University of Hawaii Press.
- Graham, Angus Charles. 1958. *Two Chinese Philosophers*. London: Lund Humphries.
- Harbsmeier, Christoph. 1993. Conceptions of knowledge in ancient China. *Pages 11–30 of: Lenk, Hans,*

- & Paul, Gregor (eds), *Epistemological issues in classical Chinese philosophy*. State University of New York Press Albany, NY.
- 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. 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, Wang Yangming, and Mengzi's contributions to understanding moral perception. *Dao*, **10**(3), 273.
- Johnston, Ian, & Ping, Wang. 2012. *Daxue and Zhongyong: Bilingual Edition*. Chinese University Press.
- Lederman, Harvey. 2020a (January). *Perception, Know-How and Genuine Knowledge in Wang Yangming*. Unpublished MS.
- Lederman, Harvey. 2020b (October). *What is the "unity" in the "unity of knowledge and action"?* Unpublished MS.
- 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.
- 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. 1996. *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.
- Shun, Kwong-loi. 2011. Wang Yang-ming on Self-Cultivation in the *Daxue*. *Journal of Chinese Philosophy*, **38**(s1), 96–113.
- Shun, Kwong-loi. 2018. On the Idea of 'No Self'. *Proceedings and Addresses of the American Philosophical Associations*, **92**, 78–107.
- Tang, Chun-i. 1973. The Criticisms of Wang Yang-ming's teachings as raised by his contemporaries. *Journal of Chinese Philosophy*, **23**(1/2).
- Tiwald, J., & Van Norden, B.W. 2014. *Readings in Later Chinese Philosophy*. Hackett Publishing Company, Incorporated.
- Tu, Wei-ming. 1976. *Neo-Confucian Thought in Action: Wang Yang-Ming's Youth*. Berkeley: University of California Press.
- 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.
- Zhu, Xi (朱熹) (ed). 1983. *Collected Commentaries on the Four Books (四書章句集註)*. 北京: 中華書局.