

[Author's copy of a paper forthcoming in *The Journal of Moral Education*. Please refer to the published version in citations.]

## CAN WE MEASURE PRACTICAL WISDOM?

Jason Swartwood

Department of Philosophy

Saint Paul College

235 Marshall Ave,

Saint Paul, MN 55102

[jason.swartwood@saintpaul.edu](mailto:jason.swartwood@saintpaul.edu)

**Abstract:** Wisdom, long a topic of interest to moral philosophers, is increasingly the focus of social science research. Philosophers have historically been concerned to develop a rationally defensible account of the nature of wisdom and its role in the moral life, often inspired in various ways by virtue theoretical accounts of practical wisdom (*phronesis*). Wisdom scientists seek to, among other things, define wisdom and its components so that we can measure them. Are the measures used by wisdom scientists actually measuring what philosophers have in mind when they discuss practical wisdom? I argue that they are not. Contemporary measures of wisdom and its components may pick out some necessary prerequisites of practical wisdom, but they do not measure a philosophically plausible practical wisdom or its components. After explaining the argument and defending it against objections, I consider its implications. Should wisdom scientists ignore the philosophical conception of practical wisdom in favor of other conceptions, revise their methods to try to measure it, or continue the interdisciplinary study of practical wisdom without expecting to measure it? I make a preliminary argument for the third option.

**Keywords:** wisdom, practical wisdom, wisdom science, phronesis, Berlin Wisdom Paradigm, Three-Dimensional Wisdom Scale, Situated Wise Reasoning Scale.

**Bio:** Jason Swartwood is instructor of philosophy at Saint Paul College. His main research interest is developing and examining empirically-informed accounts of practical reason. He has published work on practical wisdom, methodology in practical ethics, and pedagogy in moral philosophy. He is also interested in developing strategies for successfully imparting moral reasoning skills in ethics courses and is currently at work on a practical ethics textbook (co-authored with Ian Stoner), titled *Doing Practical Ethics*, under contract with Oxford University Press.

## 1. INTRODUCTION

Wisdom, long a topic of interest to moral philosophers, is increasingly the focus of social science research.<sup>1</sup> Philosophers have historically been concerned to develop a rationally defensible account of the nature of practical wisdom (*phronesis*) and its role in the moral life. Wisdom scientists seek to, among other things, define wisdom and its components in order to measure them. In many cases, wisdom scientists reference philosophical accounts of wisdom when explaining or developing their definitions, but they make no explicit claim that their definitions or corresponding measures pick out the type of wisdom philosophers are interested in. In other cases, wisdom scientists do explicitly claim they have the same target as philosophers.<sup>2</sup> Given the promise of interdisciplinary collaboration on such an important topic, we should wonder: *are* wisdom scientists measuring practical wisdom, as it is understood by philosophers?

I will argue that they are not. Whatever prominent measures of wisdom and its components pick out, they do not measure a philosophically plausible practical wisdom. More specifically, they may be *measures of some practical wisdom-relevant characteristics* – characteristics that provide an incomplete and underspecified set of some important necessary conditions of practical wisdom. They nevertheless do not qualify as *measures of practical wisdom or its components*. Just as measures of an accurate grasp of chess rules (how you can move a pawn, etc.) would not qualify as measures of chess expertise or its components, prominent wisdom measures do not qualify as measures of practical wisdom or the knowledge, reasoning processes, or personal characteristics it is composed of.

I begin in section 2 by using examples from the empirical literature on expertise and expert performance to isolate a necessary condition for a valid measure of practical wisdom and its components. In section 3, I explain why prominent wisdom measures do not satisfy this condition. In section 4, I review the argument and its implications. Finally, in section 5, I discuss how wisdom scientists might respond. If the argument is a good one, wisdom scientists could respond with optimism that the deficiencies identified by the argument can be overcome. They might also try to distance themselves from the philosophical conception of practical wisdom and suggest they are examining different but equally interesting definitions of wisdom. I explain why I find both of these responses unsatisfying and suggest that philosophers and scientists interested in the interdisciplinary study of practical wisdom should shift their focus away from measuring it.

## 2. A NECESSARY CONDITION FOR MEASURES OF PRACTICAL WISDOM OR ITS COMPONENTS

On the philosophical conception I'm concerned with, practical wisdom is the understanding that enables a person to make reliably excellent decisions about how they<sup>3</sup> ought to live. I'll explain the central features of this conception in a later section. For now, the main point is that practical wisdom is a high-level achievement. The practically wise person reliably excels at decisions about how to live. This excellence is challenging to achieve, even if we don't look for perfection. In this section, I'll argue that examining how we can tell if measures of other high-level achievements (such as expertise in chess and teaching) measure what they're supposed to, we can learn how to tell if a wisdom measure actually measures practical wisdom or its components.

## **2.1 Expertise and its components can be measured only if measures are specified relative to success conditions for the domain**

An expert is someone who has a reliably superior grasp of what to do across a suitably broad and challenging range of situations characteristic of their domain.<sup>4</sup> An expert chess player, for instance, reliably grasps how to make moves that win games against suitably challenging opponents. Even if someone is *viewed* as an expert (because of their experience or reputation), they are not *actually* an expert unless they possess, exhibit, or are capable of exhibiting this kind of reliably superior grasp of what to do. A person's claim to being a chess expert would be undermined if they were found to have won games only against inexperienced six year-olds.

Using this insight about the connection between expertise and expert performance, an influential body of research has shown that there are some domains where so-called 'experts' do not perform better than novices or people using simple decision-aids (Ericsson, 2018, p. 14). The research has also identified measures that successfully pick out expertise and the extent to which people have it, along with other measures in those same domains that fail to do so. Attention to these cases provides general lessons about when measures of high-level achievements, including wisdom, succeed or fail in measuring what they're supposed to.

### *2.1.1 Examples: valid and invalid measures*

One prominent line of research on expertise and expert performance has focused heavily on chess. A chess expert is someone who reliably wins (or is capable of winning) games. To measure chess expertise, we need to identify the characteristics that contribute to or correlate with winning games against a range of opponents in a representative set of circumstances. Chess has been a focus of much research in part because there are sophisticated ways to do this. For instance, the Elo scale provides a numerical value of chess expertise calculated based upon performance in tournaments (Elo, 1986; Gobet & Charness, 2018). The Elo scale allows us to measure, to a fine degree, how much chess expertise someone has.

Because we can measure chess expertise in this way, we can also identify what other characteristics correlate with chess expertise. Is an expert chess player distinguished from a novice (or less expert) player by their possession of specific knowledge, certain processes of reasoning or decision-making, memory capabilities, or other personal characteristics? By seeing which features do or do not correlate with expert performance (quantified by the Elo scale), research has shed light on this question.

Although we might suspect expert players think farther ahead than novices about the consequences of their moves, early research showed that expert players did not exhibit this ability to a greater degree than novice players (Gobet & Charness, 2018). We might also suspect that expert chess players have greater memory recall than less expert players, but research has shown at best weak correlations between expert performance and success in recall tasks (Gobet & Charness, 2018; Moxley & Charness, 2013).

A different measure has been found to correlate significantly with Elo rankings. The "best move task" asks participants to select the best move for a particular board arrangement. What counts as a best move is defined by studying the games of experts, who were in turn identified by measures of expert performance (such as Elo rankings). Research has found that successful performance on the task

correlates significantly with national and international chess rankings (Moxley & Charness, 2013). Indeed, it correlates much more significantly than other measures, such as measures of memory recall, thus making the best move task a better measure of chess expertise (ibid). We know that measuring a person's performance on the best move task more accurately measures their level of skill than memory recall tasks, because the former correlates more significantly with expert performance.

Teaching is another area where research has identified valid and invalid measures. An expert teacher is someone who reliably grasps how to inculcate knowledge, skills, or other valuable components of personal growth. To measure teaching expertise, we need to identify characteristics that contribute to, or correlate with, reliably inculcating knowledge or skill in suitably receptive students in a given population.<sup>5</sup>

One challenge in measuring expertise in teaching is defining the goals of teaching. Whereas some teachers may be better at helping students pass standardized tests, others may be better at teaching important life skills and habits (such as grit) (Stigler & Miller, 2018, p. 431). Unsurprisingly, much research has focused on which teachers help students pass standardized tests. The MET (Measures of Effective Teaching) study developed a measure of teaching expertise and tested it using a random assignment experiment. Researchers first developed a measure of teaching expertise calculated based upon various teacher and student characteristics (including past student performance on standardized tests, whether students receive free or reduced lunch, etc.) combined with observations from video-taped lessons and feedback from student evaluations of teaching (Kane, McCaffrey, Miller, & Staiger, 2013, p. 6). Then, they followed more than 3,000 teachers in seven districts in different U.S. states, and students in each school were randomly assigned teachers. The researchers wanted to know: would the measure accurately predict student achievement when students were assigned randomly to a teacher? The results indicated that the measures accurately predicted the degree of achievement a teacher's students accomplished on state tests and specific Math and English assessments (Kane et al., 2013, p. 2). These results indicate that there are some teachers who are significantly better than their peers at helping their students pass standardized tests, and we have measures that can help identify these teachers. So, while there is room to doubt whether getting students to pass standardized tests is actually the goal of teaching expertise, expertise at helping students pass standardized tests is something we can measure.

Other measures of teaching expertise have not fared as well. Years of experience teaching do not correlate well with student learning outcomes, so experience is not a good measure of teaching expertise (Stigler & Miller, 2018). The National Board for Professional Teaching Standards (NBPTS) certifies teachers based upon assessment of a teaching portfolio and examination. Research indicates that NBPTS correlates somewhat with student outcomes (performance on writing tasks or math tests, etc.), but the correlation has weakened in recent years as more teachers have sought certification (ibid).

In the MET study researchers tried to apply purely observational measures to video performances of classroom teaching. To do so, trained raters watched videos of classroom performance and rated the degree to which teachers exhibited best practices such as used an investigation/problem-based approach, engaged students, managed student behavior, showed good time management, gave quality feedback, and so on (Kane et al., 2013, p. 20). Importantly, the resulting measure failed to correlate significantly with gains in student achievement in the way the primary MET measure did (Stigler & Miller, 2018, p. 438).

Studies on student evaluations of teaching (SET) have shown that they also fail to measure teaching expertise, because they favor teachers based upon pedagogically-irrelevant features (attractiveness, gender or race of the teacher, easiness of the workload, etc.) and fail to significantly correlate with learning outcomes (Boring, Ottoboni, & Stark, 2016, p. 10; Greenwald & Gillmore, 1997; Gump, 2007; Hessler et al., 2018).<sup>6</sup>

*2.1.2 The best explanation of the examples is that a clear and representative account of success conditions is necessary for measuring expertise and its components*

We can derive from these examples a general lesson about what is required to show that a measure of high achievement in a domain is valid – that it actually measures what it is supposed to. The reason we can identify which measures accurately pick out expertise in chess and teaching and which do not is that we have identified clear and representative success conditions in those domains.

A chess expert is someone who exhibits (or is capable of exhibiting) expert performance, which means reliably winning suitably tough games. We can measure who is a chess expert because we have clear and precise ways of telling who has won games and who has not. We can tell that memory recall tasks and experience are worse measures of chess expertise than the best move task precisely because we can show that the former do not correlate as significantly with successfully achieving the ends of the domain (winning games) as the latter does.

An expert teacher is someone who exhibits (or is capable of exhibiting) expert performance, which means reliably increasing students' learning gains. We can measure who is an expert teacher to the extent that we have clear and precise ways of telling when a student has made learning gains and when they have not. We can tell that observing general 'best practices' is not a good measure of expertise and that the MET is a good measure precisely because we can show that the former does not correlate significantly with successfully achieving the ends of the domain (student learning gains) while the latter does.

Indeed, without clear and representative success conditions for performance in a domain, we would not be able to tell if a measure of expertise in that domain is valid. Consider a hypothetical game, Mess, which uses the same pieces and board as chess. Suppose we didn't have a clear way to tell when someone had won a game of Mess. Suppose there was no clearly identifiable end point of games and winning was determined not by the eventual presence of a specific type of board arrangement but by the degree to which the player's moves all fit together into a coherent and aesthetically pleasing strategy. Suppose also that there was vast disagreement about what counted as a better strategy; there was agreement that not all are equally good, but the details about which moves should be made respective to others were not agreed upon.

The lack of a clear and representative account of success would prevent us from identifying a valid measure of Mess expertise. We wouldn't be able to measure who is an expert or which particular characteristics (such as level of experience, or an ability to articulate general rules for good play, an ability to remember board positions, etc.) are indicative of expertise, because we lack the ability to tell whether someone with those characteristics is more likely to win than someone without them. Importantly, if the characteristics of moves that contribute to winning are complex enough, it would be challenging to measure Mess expertise even if we had a plausible way to rationally resolve disagreement about general characteristics of a win.

Clear and representative success conditions are thus necessary for a measure of expertise. They are also necessary for measures of its components.

By a *component of expertise*, I mean a characteristic or feature possessed or exhibited by experts that distinguishes them from novices or those with significantly less expertise. On this definition, not all necessary conditions of expertise are components of expertise. Consciousness is necessary for human chess expertise (it would be hard for a human to win games without being conscious) but we would not call it a component of expertise, because it is shared by novices. Knowing the rules of chess (e.g. what moves the rules allow you to make in a particular board arrangement) is similar. Characteristics like these that are necessary conditions of expertise but that do not tend to distinguish novices from experts could be called *expertise-relevant characteristics*. On the other hand, we can imagine that expert chess players possess or exhibit knowledge, processes of decision-making, or other characteristics that novices (and those who are less expert) tend to lack. If there are such characteristics, then they would qualify as components of chess expertise.

Importantly, we can only measure these components if we specify them in a way we can verify significantly correlates with successfully achieving the ends of the domain. For instance, we can say that the ability to recognize patterns quickly is a component of chess expertise to the extent that we can show this ability correlates with winning games. But, this ability can be defined and operationalized at various levels of generality, not all of which correlate significantly with winning. A pattern recognition ability may be a component of chess expertise. But, to measure this component, we need to specify it in a way that distinguishes novices who quickly recognize patterns (but still don't win) from experts who see relevant patterns that contribute to winning.

The research on teaching provides an additional illustration of this point. James Stigler and Kevin Miller argue that observational best practices measures fail to pick out expert teachers in part because of

... the contextual nature of teaching. A move that might be best in one situation might be precisely the wrong move to make in a different situation. For example, a critical remark to one student may be just what he needs to engage him in digging deeper on a problem. But the same remark may come as a crushing blow to a different student, who needs encouragement.

... With goals in mind, teachers must constantly read the situation, monitor progress, and make necessary adjustments.

This analysis leads us to reject the idea that expertise in teaching can be defined in terms of decontextualized "best practices." Our view is that correlations between teacher actions and student learning are low not because we haven't yet identified the right set of best practices, but because teaching itself is contextual, meaning that such correlations will always be low. (Stigler & Miller, 2018, p. 438)

I suggest we can infer from this the following lesson: in *high-context domains* (domains where the factors that constitute or contribute to achieving domain success are many, varied, and interact in complex ways), measures of components of expertise must be specified relative to successful performance. It is true that 'gives quality feedback' is a best practice for teachers, so we can expect that the ability to do so is a component of expert teaching. However, to measure the ability to give quality feedback requires specifying what counts as quality feedback in particular situations. Since the quality is

determined by the extent to which it contributes to student learning, and since the factors impacting whether it will do that are many and complexly related, this is a formidable task. Indeed, Stigler and Miller suggest that the project of specifying ‘best practices’ measures so that they adequately distinguish experts is unlikely to succeed. It is reasonable to expect that attempts to use generally described characteristics to measure high-level achievements in other high-context domains will fare similarly. Such measures deserve our confidence only if we can verify they significantly correlate with successful performance in ways that distinguish people with significantly different levels of expertise.

Using some representative examples, I have argued for the following claims:

- We can be confident a measure of expertise or its components is valid (i.e. measures what it is supposed to measure) only if it is specified relative to clear and representative success conditions for the domain.<sup>7</sup>
- Measures of generally described characteristics of expertise will be unlikely to be valid measures of expertise or its components to the extent that the domain is high-context.

## **2.2 Practical wisdom is (or is analogous to) expertise**

Specifying our measures relative to success conditions is thus a necessary condition of identifying valid measures of expertise and its components. Because practical wisdom is a form of expertise in the relevant sense, the same condition applies to measures of practical wisdom and its components.

The sense of ‘expertise’ I’m relying upon is not a highly theorized one. By expertise, I just mean a high-level achievement tied to superior performance in an area. Chess expertise is a high-level achievement in this sense, because a chess expert is someone who demonstrates (or is capable of demonstrating) reliably superior achievement in chess matches. Teaching expertise is a high-level achievement in this sense, for similar reasons.

The same could be said of practical wisdom: it is a high-level achievement, because a practically wise person has the understanding required to make reliably good decisions about they ought to live. So, while psychologists (Baltes & Staudinger, 2000) and philosophers (Annas, 2011; Swartwood, 2013) have defended definitions of wisdom that describe it as a form of expertise, my argument here need not assume any of those accounts. Because practical wisdom is a high-level achievement in the same way that (for example) chess expertise and teaching expertise are, the same necessary condition applies to measures of practical wisdom and its components.

## **2.3 Success in the domain of practical wisdom is success in particular all-things-considered decisions**

In order to measure practical wisdom or its components, our measures will need to be specified relative to success conditions. But what are the success conditions for practical wisdom? On a minimal philosophical account, practical wisdom is a grasp (1) of what one ought to do, (2) all-things-considered, (3) in particular situations. Examining this account reveals necessary conditions for valid measures of practical wisdom.

I am not claiming this account (which I’ll call “the minimal philosophical conception”) is universally accepted by philosophers (or anyone else). I merely claim that the conception is rationally compelling and of significant historical and contemporary interest. To see why, consider the three features of the conception in turn.

### 2.3.1 *The domain of wisdom: decisions about what one ought to do*

On the account I'm describing, practical wisdom is a grasp of how one *ought* to live and conduct oneself.<sup>8</sup> We all need to grapple with questions about what we ought to do. Should I take the job I really find rewarding when the commute will reduce time with my family? Should I seek refugee status to escape violence in my home country when it could result in being separated from my children? How should I take care of my family with the limited resources I have? What should I do with my expendable income? How should I respond to a racist or sexist slight at work? Should I tell my sister my concerns about the suitability of her fiancé? When and how should I challenge political or social views I think are misguided? (Which views are actually misguided?) What's the best way to deal with a challenging co-worker, student, boss, spouse, or child? Which of my reactions to stressful situations are good for me, which are not, and how can I improve them? When should I do good for others even when it's not necessarily best for me?

Some decisions like these are relatively easy (should I yell at my partner for calmly raising an important concern about our finances?). Many others are very challenging, even if we sometimes don't recognize them as such. To make good decisions about what we ought to do, we can't look merely at what people tend to do, what others think we should do, what we want to do, or even what's considered wise by our society or group. An authoritarian autocrat or con artist are not made wise simply because they believe they are wise. Sexual harassment, enslavement, disregard of people with disabilities, and corporal punishment for children are not made wise because they are endorsed by your culture, or even the majority of people or cultures.

On the minimal philosophical conception, having practical wisdom means having a reliably good grasp of what one ought to do. In philosophical terms, practical wisdom is a prescriptive ideal because it deals with how we *ought* to be and conduct ourselves, not simply what or whom people think is wise or how people tend actually to be or behave (Swartwood & Tiberius, 2019).<sup>9</sup>

### 2.3.2 *The domain of wisdom: all-things-considered decisions*

Suppose you find that your spouse has committed a crime: they worked out a deal with your child's college admissions test coach to have someone raise your child's score artificially. Your child just received an acceptance packet from her college of choice and is thrilled, and your spouse has no clue you're aware of the deal. What should you do? If you asked a lawyer for advice, their guidance might be aimed at keeping you out of legal trouble. They would tell you what you ought to do given that specific goal. But living wisely requires evaluating and balancing the various goals one might have. In this case, should you tell your child the truth, or should you tell your spouse you know? Should you stick by your spouse and do what you can to protect them from harm? Should you do what promotes your child's well-being? What's that? Or do considerations of fairness imply you should somehow bring your spouse's misdeeds to light? If so, what's the best way to do that?

What makes decisions like this challenging is that they are not simply a matter of figuring out what action will achieve a specific goal or outcome. Instead, they require us to evaluate what we ought to do given the various different goals we might take up, which requires evaluating, specifying, and balancing those goals.<sup>10</sup> Put simply, we need to decide what we ought to do *all-things-considered*.



This is why practical wisdom, on the minimal philosophical conception, is the knowledge or understanding required to make reliably excellent decisions about what one ought to do all-things-considered (Hursthouse, 1999, pp. 59–61). A person with practical wisdom grasps what they ought to do, all-things-considered, even in situations that would befuddle the rest of us.

### 2.3.3 *The domain of wisdom: decisions in particular situations*

Practical wisdom is thus a grasp of what one ought to do all-things-considered, and a person is practically wise to the extent that they have this grasp. But, how can we tell if someone has this grasp, as opposed to a grasp that is not indicative of practical wisdom?

A plausible answer is that wisdom is the grasp that is conducive to living a good human life (Aristotle, 1999; Grimm, 2015; Hursthouse, 1999; Kraut, 1993; McDowell, 1979, 1996). Most Virtue Theorists, inspired by Aristotle, would insist that a good human life is not simply the life that produces the most pleasure or subjective or psychological well-being (Foot, 2003; Hursthouse, 1999; Kraut, 2009; Nussbaum, 2001, 2008, Chapter 9). While there is disagreement about which alternative account of a good human life is most plausible, none are expected to float free of what ought to be done in particular situations. Practical wisdom is the understanding that contributes to living a good human life, and living a good human life is a matter of exercising a reliably good grasp of what one ought to do, all-things-considered, in particular situations. Just as a person would not count as a chess expert if we found they did not, in fact, win many games, a person would not count as having practical wisdom if they did not have a reliably good grasp of what to do in particular cases.<sup>11</sup>

Importantly, the domain of wisdom is extremely high-context. There is no simple way to describe what counts as a successful grasp of what one ought to do, all-things-considered, in particular situations. Consider these examples:

**Santa:** In the United States, it is not uncommon for parents to tell their children that Christmas presents are delivered by Santa Claus – an overweight, jolly, bearded man who wears red, lives in the North Pole with elves, and drives a sleigh led by flying reindeer. Cassandra has been approached by a gaggle of her nieces and nephews, who have heard from one of the older kids that Santa Claus is not real. What should she tell them?

**Undiagnosed:** Though she's not a psychologist, Maryam has grown familiar with ASD (Autism Spectrum Disorder) through her work as a teacher. This familiarity has led to a difficult decision. Maryam's friend Jules has recently shared a number of challenges she's encountered with her three-year-old son. Jules' son is very rigid and has tantrums when things are not just so, he only eats four different foods, he does not seem interested in engaging with other people socially, and he is obsessively focused on airplanes. Jules is clearly distressed by her son's behavior, which is causing stress at home between herself and her husband. Maryam knows that Jules and her husband have expressed skepticism about psychologists, whom they think are responsible for pathologizing kids. 'Why can't they just let kids be kids?' they often say. Jules has made it clear she's not interested in Maryam's views about what is happening with her son, even though she's very aware that Maryam has experience helping kids of all sorts. Maryam wonders if she should tell Jules to have her son assessed for ASD, just in case. Getting help early could lead to great gains for her son. But, she worries about how Jules will react if she mentions things. What should she do?

**Caretaker Fib:** Xiong has a close relative who suffers from schizophrenia. The relative has extreme paranoia and is deeply concerned about being manipulated against their will, and he sometimes has a

hard time distinguishing reality from fantasy. Xiong sometimes helps care for him. The relative is not able to live independently but is intelligent and, with the help of medication, capable of forming some meaningful bonds with others. For reasons he hasn't fully explained, he refuses to take his anti-psychotic medication when Xiong brings it to him. Without the medication, his hallucinations and distress grow worse, though the medication does also have a sedative effect Xiong's relative seems to dislike. Xiong has discovered that lying to his relative – telling him he's receiving Tylenol rather than the anti-psychotic – makes him more willing to take the medicine. Xiong is unsure how else to get him to take his medication. What should Xiong do?

**Benevolent Scapegoating:** Raheem is a student at a small liberal arts college in the Midwestern United States. Raheem and his friends have been frustrated by the lack of productive discussions about racism in the community, which is predominantly white but has a still sizable group of racial and ethnic minorities. A recent incident has inflamed the tensions but seems to be bringing the issue to the forefront in a productive way. A racist note was found placed on someone's car in a college parking lot, and this has sparked an investigation by the college administration. Members of the college and the community at large are having more discussions about the problem of racism. Raheem has felt hopeful that these conversations, though challenging and painful, could lead to progress. To his surprise, Raheem has found out that one of his acquaintances at school, who also shares his frustration with the lack of progress, actually forged the racist note to try to start a conversation about racism. Raheem sees that this deceit may result in some good for the community, but there are also some potential downsides for individuals who are being investigated. Plus, the cause would potentially be harmed if the deceit is revealed. What should Raheem do with this information?<sup>12</sup>

These scenarios are thinly described, and in the real-world various further details might be relevant. And, one person is unlikely to encounter all of these. But I suspect that, on reflection, you'll agree to the following three points about the cases.

First, these cases are structurally similar to the kinds of situations people routinely face in their lives, and a person is wise to the extent that they respond well in particular situations such as these. These cases involve common patterns of conflict between just a few of the commitments that are part of any good human life (Nussbaum, 2008). Regardless of culture or circumstance, people aiming to live a good life will have to figure out how to balance truth-telling with (among other things) commitments to loyalty, fairness, respect for autonomy, and compassion.

Second, there are better and worse ways to conduct oneself in such situations, even if it's not always clear to us which ways are which. We're compelled to search for the best approach to situations like these, and it's hard to believe this urge is fundamentally misguided.

Third, there is no simple way to describe the end of wise decisions in cases like these, because there is no simple way to balance commitments to the various values at stake. In the examples above, each scenario highlights the way one particular value (truth-telling) interacts with our other commitments (such as promoting others' wellbeing, promoting our own well-being, promoting justice and fairness, and so on). It might be tempting to say that the wise decision in each of these cases is the one that produces good for society.<sup>13</sup> But how should we specify that good so that we're given specific guidance in all the particular cases? The maxim 'wise conduct is conduct that produces the best outcome for society' might seem tempting, but it is both underspecified<sup>14</sup> and gives implausible and oversimplifying guidance in at least some of the cases (Benevolent Scapegoating and Undiagnosed, for example). I suspect you'll find that other simple attempts to describe the features of wise conduct in particular

situations fare the same. This becomes even more apparent when we consider additional cases that focus on things other than truth-telling. Although there are better and worse ways to deal with particular all-things-considered decisions, there is no simple set of rules we could use to capture what ought to be done, all-things-considered, in all situations a person encounters in their life (Hursthouse, 1999, p. 56).<sup>15</sup>

#### **2.4 Implication: we can measure practical wisdom or its components only if our measures are specified relative to success in particular all-things-considered decisions**

I have described a minimal philosophical conception of practical wisdom, and the argument so far implies a necessary condition for measuring practical wisdom, so conceived.

Practical wisdom is a grasp of what one ought to do all-things-considered, and the success conditions for practical wisdom and its components are success in particular all-things-considered decisions. Therefore, we can be confident that we are measuring practical wisdom only if the measure is specified relative to success in decisions about what one ought to do, all-things-considered, in particular situations. As I have emphasized, there are good philosophical reasons to say that a simple and comprehensive account of those success conditions is not possible.

Although the conception of practical wisdom I've described here has its origins in philosophy, I've tried to stress why it should not be appealing only to philosophers. Research on implicit theories of wisdom indicates that "people have associated wisdom with a repertoire of practical strategies, competencies, and skills that are judiciously applied in problem-solving, decision-making, and life management contexts" (Weststrate, Bluck, & Glück, 2019, p. 105). The thought that wisdom enables good decision-making in challenging situations also seems to be an assumption behind many explicit theories of wisdom developed by wisdom scientists (Ardelt, 2004, p. 275; Baltes & Staudinger, 2000; Sternberg, 1998, p. 355). One recent interdisciplinary model of wisdom explicitly mentions all three of the features of the minimal philosophical conception I have defended (Darnell, Gulliford, Kristjánsson, & Paris, 2019, p. 13).<sup>16</sup> Conceiving practical wisdom as a grasp of what one ought to do, all-things-considered, in particular situations makes coherent the compelling idea that wisdom enables good decision-making in the concrete circumstances of our lives.

In the context of the argument of this paper, the conception I've defended is also attractive because it is neutral on many of the substantive theoretical disagreements philosophers and psychologists have about wisdom. For instance:

- **Wisdom as a state vs. trait:** Wisdom scientists have disagreed about whether wisdom is a trait (a stable and invariable individual disposition), a state (a situation-specific expressions of wise behavior), or an integration of the two (Grossmann, 2017; Grossmann, Dorfman, & Oakes, 2019; Grossmann, Kung, & Santos, 2019). As I will emphasize, my goal is to show that contemporary wisdom measures do not measure the minimal philosophical conception of practical wisdom, regardless of whether we view practical wisdom as a state, a trait, or an integration of the two.
- **Wisdom, affect, and motivation:** Philosophers disagree about whether a practically wise person's grasp of what to do will necessarily include affect that motivates them to take action. On one influential picture, a person with practical wisdom not only grasps what action to take but grasps the correct reasons it ought to be done and does so in a way that motivates action (Ivanhoe, 2002; McDowell, 1979; Mengzi, 2008). According to other views, practical wisdom

includes knowledge of what one ought to do but not necessarily the character virtues that motivate one to do it (Stichter, 2016; Wolf, 2007).<sup>17</sup> Both perspectives are compatible with viewing practical wisdom as a grasp of what one ought to do, all-things-considered, in particular situations.

In these and other respects,<sup>18</sup> the philosophical conception I've defended here is neutral on substantive theoretical questions about wisdom. So, while there are certainly other possible conceptions of wisdom we could adopt, the one I've described is both compelling and relies on minimal controversial assumptions.

### 3. CONTEMPORARY WISDOM MEASURES DO NOT SATISFY THIS NECESSARY CONDITION

Even if the philosophical conception of wisdom I've described is compelling, it would be hasty to assume that wisdom scientists have intended their measures to operationalize it. Still, the plausibility of the minimal philosophical conception and its overlap with the conceptions used by wisdom scientists and everyday people makes it interesting to ask: do contemporary wisdom measures actually measure practical wisdom as described in the minimal philosophical conception?

Wisdom scientists use a variety of methods to measure wisdom and its components, including performance measures and self-report tasks (Glück, 2017; U. Kunzmann, 2019; Webster, 2019). All these types of measures fail to meet the standard, outlined above, for valid measures of practical wisdom or its components.

#### 3.1 Measures of wise knowledge

Consider first a prominent measure of the knowledge component of wisdom: the Berlin Wisdom Paradigm (BWP), developed by Paul Baltes, Ursula Staudinger, and colleagues (Baltes & Staudinger, 2000; Ute Kunzmann & Baltes, 2003). According to the BWP, the knowledge component of wisdom is an "expert knowledge system concerning the fundamental pragmatics of life" (Baltes & Staudinger, 2000). This account of wisdom is intended to measure only the cognitive aspects of wisdom, not the affective or reflective aspects (Ute Kunzmann & Baltes, 2003). On this view, wise knowledge satisfies five criteria:

- Rich factual knowledge about life meaning and conduct
- Rich procedural knowledge about life meaning and conduct
- Lifespan contextualism
- Value relativism and tolerance
- Recognition and management of uncertainty

To measure the degree to which someone has wise knowledge, the Berlin Wisdom Paradigm asks people to think out loud about how they'd respond to a challenging life problem, question, or decision. Participants' responses were recorded and evaluated by trained raters who applied the five criteria.

Here's one prompt used in the measure and an example of a highly-rated response (Baltes & Staudinger, 2000, p. 136):

A 15-year-old girl wants to get married right away. What should one/she consider and do?

[Highly rated answer]: Well, on the surface, this seems like an easy problem. On average, marriage for 15-year-old girls is not a good thing. But there are situations where the average case does not fit. Perhaps in this instance, special life circumstances are involved, such that the girl has a terminal illness. Or the girl has just lost her parents. And also, this girl may live in another culture or historical period. Perhaps she was raised with a value system different from ours. In addition, one has to think about adequate ways of talking with the girl and to consider her emotional state.

For reasons that should be clear now, we have no reason to think this procedure picks out knowledge that distinguishes those with practical wisdom from those without. It is plausible to say that a wise person knows that different cultures have different values, that they will be able to explain some general features of well-lived lives, and that they know many facts about people and possible outcomes of different courses of action. But, just as a body of knowledge won't count as a component of chess expertise unless it is conducive to winning, a body of knowledge doesn't count as a component of practical wisdom unless it is conducive to reliably good all-things-considered decisions.<sup>19</sup> Since the knowledge measures are not specified relative to clear and representative success conditions for particular all-things-considered decisions, the BWP does not measure a *component of practical wisdom* (a feature that distinguishes people with practical wisdom from those without), though it likely does measure a *practical wisdom-relevant characteristic* (a characteristic that is a necessary condition for practical wisdom but does not adequately distinguish people with practical wisdom from those without).

To illustrate, the highly rated answer above may have been offered by someone who has practical wisdom. But, it might not. A clever undergraduate who has applied themselves to their cultural anthropology class might give this answer, and that indeed indicates that they can at least recognize some of the general ends that matter in situations like this. But this level of cleverness is quite different from the situation-specific grasp of how best to balance these ends (Annas, 2004; Aristotle, 1999, II. 1144a25-36; Hursthouse, 1999, p. 61). (This is not to say anything negative about the participants in these studies, any more than questioning the accuracy of an evaluation of teaching is impugning the teachers it has been used on.) Put differently, we have no more reason to think that a person who gives the highly rated answer above will make reliable good decisions in scenarios of the type discussed in section 2.3 than we do to think that a teacher who knows that giving quality feedback is an important part of teaching will actually perform as an excellent teacher.

Importantly, the problem would still remain even if the scenario were spelled out in more detail. Raters who apply the five criteria to evaluate replies to this prompt will not be able to non-arbitrarily discriminate between answers that exhibit significantly different levels of wisdom. This is because the five criteria are not specified at the level of specific all-things-considered decisions. Indeed, for the reasons discussed above, they would be bound to oversimplify matters if they were. More recent research utilizing the Berlin Paradigm has adapted the measure to focus more precisely on age effects in participants' responses (Thomas & Kunzmann, 2013), but they do also do not specify the rating criteria relative to success conditions for particular all-things-considered decisions.

### **3.2 Measures of wise personality traits**

Other wisdom measures define wisdom as a personality trait. A prominent example is sociologist Monika Ardelt's Three-Dimensional Wisdom Scale (Ardelt, 2003, 2004). Whereas the Berlin Wisdom Paradigm only attempts to measure the knowledge component of wisdom, the Three-Dimensional Wisdom Scale (3D-WS) describes wisdom as a personality trait that also includes affective and reflective dimensions. To measure wisdom, the 3D-WS compiles the results of older adults' agreement or disagreement to 39 statements, including:

- Things often go wrong for me by no fault of my own.
- A problem has little attraction for me if I don't think it has a solution.
- Ignorance is bliss.
- Before criticizing somebody, I try to imagine how I would feel if I were in their place.
- I often have not comforted another when he or she needed it.
- People make too much of the feelings and sensitivities of animals.

People who respond to these statements in identical ways might make drastically different decisions about what they ought to do in particular situations, such as those discussed above. Such differences are not ancillary to the person's degree of practical wisdom; they are directly representative of it. Because statements included in the 3D-WS are not specified by ensuring they correlate with reliably good decisions about what one ought to do in particular situations, we have no reason to think it measures the personality components of wisdom. The situation is similar to measures of teaching expertise that focus on general 'best practices' statements ('good teaching requires giving quality feedback to students'). We could be confident that these actually measure teaching expertise only if we know those measures correlate with success in the domain (student learning gains). Similarly, we can be confident the 3D-WS measures the minimal philosophical conception of practical wisdom only if we have reason to think the scoring we are using correlates with a good grasp of what one ought to do, all-things-considered, in particular situations. Because we have no reason to think that, we can't be confident the 3D-WS measures practical wisdom or its components, even if it is perhaps measuring practical wisdom-relevant characteristics or wisdom on a different conception.<sup>20</sup>

It might be objected that the 3D-WS is clearly measuring characteristics, such as compassion and perseverance, that are associated with wisdom. Doesn't this show that it is measuring wisdom or its components?

To see why it does not, consider compassion. It is surely true that a practically wise person will be disposed to feel concern for others' well-being. If this is what we mean by compassion, however, it would not qualify as a component of wisdom, because it would not necessarily distinguish a person with significant practical wisdom from someone who significantly lacks it. A compassionate response is not necessarily a wise one, because your grasp of what you ought to do, all-things-considered, can be flawed even if you have other-regarding motives. A doctor who hides a patient's terminal diagnosis from them in order to save them mental anguish may be motivated by compassion, but it's not clear this is the wise move. Concern for the well-being of a person targeted by a sexual harassment complaint may motivate an unwise excoriation of the accuser. Often, the biggest challenge is not working up concern for others but figuring out when, how, and why to express that concern. Xiong, Maryam, and Cassandra (from the scenarios described in section 2.3.3) may be motivated by concern for others, but that does not mean that whatever they do displays a good grasp of what, all-things-considered, ought to be done.

Similar things can be said about other personal characteristics. We can expect a person with practical wisdom to persevere in the face of challenges. But, if perseverance is simply understood as the state or trait of not giving up when things get tough, this will not distinguish those with significant practical wisdom from those without it. A dedicated white nationalist who tirelessly pursues his racist agenda and a passionate activist who doggedly opposes women's suffrage may have perseverance, but that clearly does not mean that they display a practically wise person's reliably good grasp of what one ought to do.

So, while a practically wise person will certainly have compassion and perseverance, these characteristics do not count as components of practical wisdom, because they do not adequately distinguish people with significantly different levels of practical wisdom. Compassion and perseverance would only count as components of wisdom if defined (respectively) as the state or trait of grasping when, how, and why to respond to threats to others' well-being or to respond to adversity, all-things-considered.<sup>21</sup>

These examples illustrate a general point. Insofar as personal characteristics are going to be components of practical wisdom, they will need to be specified relative to particular all-things-considered decisions. Even considered as a cluster, generally described characteristics such as these do not count as a component of practical wisdom. Caring about the well-being of oneself and others, about truth-telling, loyalty, and forgiveness (among many other things) are all practical wisdom-relevant characteristics, because they are necessary pre-requisites for being practically wise. Nevertheless, the general personal characteristics measured by the 3D-WS are not components of practical wisdom, as described in the minimal philosophical conception, because they do not distinguish those with a good grasp of what one ought to do in particular all-things-considered decisions from those without it.

### **3.3 Measures of wise reasoning**

A recent and influential line of research has focused on measuring wise reasoning, defined as "the use of certain types of pragmatic reasoning to navigate important challenges of social life" (Grossmann et al., 2010). Grossmann and colleagues' (2010) early measure of wise reasoning used think-aloud protocols to identify the degree that people's reflection exhibits intellectual humility (awareness of limits of one's own knowledge), consideration of other perspectives, recognition of the likelihood of change, specification of various predictions contingent on how things go, search for conflict resolution, and search for compromise. To validate the measure, researchers had wisdom researchers and counseling professionals rank the participants' responses to see how they compared with the raters' rankings. The prompts used have more detail than the ones used by the Berlin Wisdom Paradigm, sometimes dealing with interpersonal conflicts or broader social conflicts between social groups (such as tensions between citizens of Tajikistan and Kyrgyz immigrants).

Later research has sought to improve the wise reasoning approach in various ways. The use of global self-reports of wise reasoning or third-party ratings is labor-intensive, hard to apply to immediate real-life decisions, and challenging to code consistently. The Situated Wise Reasoning Scale (SWIS) addresses this by having participants recall a challenging interpersonal conflict, after which they are asked questions to guide reconstruction of the situation and then questions to assess their reasoning about the situation (Oakes, Brienza, Elnakouri, & Grossmann, 2018). This and subsequent work (Grossmann, 2017; Grossmann, Dorfman, et al., 2019) has allowed researchers to focus more on how an individual's performance on the measure varies over time and between situations.

As with the measures previously discussed, these measures of wise reasoning are not specified relative to success conditions for particular all-things-considered decisions. Indeed, Oakes, Brienza, Elnakouri, and Grossmann highlight this in a recent review. Using a situation from a letter to an advice columnist (in which a woman wonders how to deal with her husband's insistence on discussing contentious political matters whenever friends come over), they comment:

... what we are illustrating with this example is that the cognitive principles involved in the wise reasoning approach do not necessarily suggest a single solution or desired outcome. Instead, they afford a metacognitive framework for working through the contingencies and elements playing a role in a given situation, promoting a bigger picture view – and thereby, more accurate understanding – of the situation. (Oakes et al., 2018, p. 208)

It is true the situations making up one life will differ in various subtle ways from those of another, and we should not assume that there is always only one good decision that could be made in the knotty situations we find ourselves in. The philosophical conception of practical wisdom does not deny that. However, the processes utilized in the wise reasoning approach are not described at the level of specificity required to distinguish people with significantly better or worse grasps of what one ought to do, all-things-considered. Although the measure may identify practical wisdom-relevant characteristics of wise reasoning, it does not measure the reasoning component of practical wisdom. Just as we have reason to think a process of decision-making in chess is part of chess expertise only if we can verify it distinguishes those who succeed in the domain (win games) from those who don't, we don't have reason to think the generally described reasoning processes used in the wise reasoning approach are part of practical wisdom unless we can verify that they distinguish those who make reliably good all-things-considered decisions from those who don't. Indeed, if generally described characteristics of actions or people are often unlikely to significantly distinguish experts from non-experts in high-context domains, it is unclear why we should expect generally described reasoning processes to fare differently. At the very least, we have no reason to be confident these process measures adequately distinguish people with significantly different levels of practical wisdom.

To illustrate, consider open-minded reasoning. If we view open-mindedness as seeking out other perspectives when deciding what to do, then open-minded reasoning does not necessarily lead to good all-things-considered decisions. Huckleberry Finn, considering whether to turn in his companion Jim, who is attempting to escape enslavement, may take a variety of perspectives into account. But, if the result is that he accepts the idea that Jim ought to be turned in because he is considered someone's property, we have good reason to doubt Huck's grasp of what ought to be done, all-things-considered. Open-minded participants in discussions of the morality or legality of abortion (or debates about what to do in the scenarios described in section 2.3.3) may come to conflicting positions, but that does not imply those decisions are equally plausible or wise. Open-minded reasoning, then, even if described as a state or a "distribution of situation-specific" (Grossmann, 2017 p. 245) expressions of taking other perspectives seriously, would not count as a component of practical wisdom, because we cannot be confident it will distinguish people with significantly better or worse grasps of what one ought to do, all-things-considered, in particular situations.

Similar points apply to other aspects of reasoning, even if taken in concert. For instance, someone considering what to do in Undiagnosed (from section 2.3) may try to decide by reflecting on the limits of their own knowledge, on the uncertainties and contingencies of the situation, on others' perspectives



on the situation, and on how to reconcile those perspectives. Still, it's unclear why we should be confident these generally described processes will reliably result in a good decision about what one ought to do, all-things-considered. When deciding how to respond to her friend's ignorance of the possible causes of her son's challenges, the biggest challenge for Maryam is not to reflect on others' perspectives and try to integrate them; the biggest challenge is to do so in a way that identifies which aspects of whose experiences (her friend's, the child's, etc.) actually matter and what that says about what she should do. The biggest challenge is not seeking a compromise but deciding whether and in what ways that actually matters. (Does appeasing her friend's aversion to psychologists really matter given the consequences for the child?) The biggest challenge is not to specify what will happen given various different contingencies but instead to determine whether and how those contingencies matter and how one ought to respond to them. And so on. On the philosophical conception I've described, we only have reason to think we're measuring the reasoning component of practical wisdom if we have reason to think the measure reliably distinguishes those with a good grasp of what they ought to do, all-things-considered, from those without. Since we have no reason to think this is the case with the wise reasoning approach, we cannot be confident it measures the reasoning component of practical wisdom.

Perhaps there are definitions of wisdom according to which whatever decision you end up making, for whatever reason, counts as wise as long as it exemplifies the general processes described in the wise reasoning approach. On the minimal philosophical conception, however, these reasoning processes count as components of wisdom only if they will tend to distinguish people with significantly better or worse grasps of what one ought to do, all-things-considered, in particular situations.

Compare teaching. It is true that an expert teacher will likely be disposed to take other perspectives seriously, to consider the limits of their knowledge, and so on. But it is not a conceptual truth that doing so automatically leads to good teaching outcomes, and (given the inadequacy of many other measures of generally-described characteristics in high-context domains) it would be a mistake to assume that describing reasoning processes in this general way will adequately distinguish teachers with significantly different levels of success at achieving those outcomes.

We have even less reason to be confident that the wise reasoning approach measures the reasoning component of practical wisdom. Given the much more significant complexity of factors governing success in all-things-considered decisions, there is great room for a person using the wise reasoning processes to come to significantly flawed all-things-considered decisions. At the extreme, supervillains and clever but unwise political strategists may be accomplished at considering other perspectives, considering the limits of their knowledge, identifying contingencies, and searching for ways to reconcile different perspectives. But this surely does not imply that they have a good grasp of what actually matters and what they ought to do in particular situations.<sup>22</sup> As the case of Undiagnosed illustrates, the same danger lurks even for those who lack nefarious motives. The wise reasoning approach identifies processes that could help one achieve a broader grasp of what a situation is like, but it contains no standards for determining which features of a situation actually matter, why they matter, and what, therefore, ought to be done. These standards are precisely what would be needed to reliably distinguish people with significantly different degrees of practical wisdom. So, while I suspect the wise reasoning approach currently gives us the most promising empirical account of practical wisdom-relevant reasoning characteristics, we have no reason to believe it measures the reasoning component of practical wisdom.

The lines of research I've described are not the only attempts to measure wisdom.<sup>23</sup> Nevertheless, other measures share the problems of the prominent and representative ones I have discussed here: because they are not specified relative to clear and representative success conditions for particular all-things-considered decisions, they do not measure a philosophically plausible practical wisdom or its components.

#### 4. REVIEW OF THE ARGUMENT

I have advanced this argument:

- 1) We can be confident we're measuring a high-level achievement or its components only if the measure is specified relative to clear and representative success conditions for the domain.<sup>24</sup>
- 2) Practical wisdom is a high-level achievement.
- 3) Success in the domain of practical wisdom is success in grasping what one ought to do, all-things-considered, in particular situations.
- 4) So, we can be confident we're measuring practical wisdom or its components only if our measure is specified relative to clear and representative success conditions for decisions about what one ought to do, all-things-considered, in particular situations [by 1-3].
- 5) Contemporary measures of wisdom and its components are not specified relative to clear and representative success conditions for decisions about what one ought to do, all-things-considered, in particular situations.

So, we cannot be confident contemporary wisdom measures are measuring practical wisdom and its components [by 4+5].

Two features of the argument are worth emphasizing.

First, the argument is just claiming that we have no reason to think contemporary measures are measuring practical wisdom or its components. Perhaps there is a different conception of wisdom they're measuring, but it is not the minimal philosophical conception I've described.

Second, the argument leaves it open that contemporary wisdom measures pick out practical wisdom-relevant characteristics, even if they don't measure practical wisdom or its components. Just as 'best practice' measures pick out characteristics that are necessary for teaching expertise without actually measuring teaching expertise, the characteristics picked out by contemporary measures (concern for others' well-being, taking other perspectives seriously when reasoning, etc.) are likely necessary for practical wisdom. This is useful, because it can help us examine relationships between some of the necessary pre-requisites for practical wisdom, even if we are not measuring practical wisdom or its components in a way that adequately distinguishes people with significantly different degrees of practical wisdom. For the reasons I have explained, however, contemporary wisdom measures are not measuring components of the minimal philosophical conception of practical wisdom.

#### 5. IMPLICATIONS

##### 5.1 Three responses: distancing, optimism, or a change of focus

If there are no good objections to the argument, then we are left with at least three responses. One response would be to distance wisdom science from the minimal philosophical conception of practical

wisdom defended here, insisting that existing wisdom measures at least pick out an important and interesting conception of wisdom, even if not the philosophical one I've described. Another response would be optimism about the prospect of new methods that would allow us to measure the minimal philosophical conception. A third option is to focus interdisciplinary research on filling out the philosophically plausible conception of practical wisdom but give up on attempting to measure practical wisdom or its components. To start a discussion of these options, I'll gesture here at why the last seems to me most fruitful.

## **5.2 Against distancing**

If the kind of wisdom picked out by contemporary wisdom measures is different from that picked out by the minimal philosophical conception, it would be interesting to think further about whether and how these distinct senses of wisdom are related and about their comparative value as ideals. But I think it would be regrettable if wisdom science gave up altogether on the minimal philosophical conception. Given the ways in which the minimal philosophical conception makes coherent the connection between wisdom and performance in life decisions – a connection seemingly assumed in the views of 'the folk' and of many researchers – it would be a shame to discard it prematurely.

Perhaps the argument here could be dismissed if it set too high of a standard – if it led us to expect an inordinately high level of discrimination from our measures.<sup>25</sup> If a measure of chess expertise allowed us to measure all but the finest distinctions between ultra-high performers, those limitations would not undermine its claim to measure chess expertise. If a measure of practical wisdom enabled us to measure all but the finest distinctions between (for example) the top 1% of the wisest among us, it would still be reasonable to say it measured practical wisdom. But, that is not the situation with contemporary wisdom measures. The argument here implies that contemporary wisdom measures do not distinguish between people with significantly different levels of practical wisdom (as conceived on the minimal philosophical conception). To distinguish between a person with a low or modest level of practical wisdom from someone with a high level, we'd need to discriminate between the quality of different ways of responding to particular all-things-considered decisions about what one ought to do. Since contemporary wisdom measures do not do this, they are more akin to best practices measures of teaching or a measure of chess play that focused on whether a player grasps the legal moves the rules allow them to make on a board arrangement. Just as those measures would perhaps measure expertise-relevant characteristics but not expertise, contemporary wisdom measures at best measure practical wisdom-relevant characteristics but not practical wisdom.

## **5.3 Against optimism**

If the minimal philosophical conception should not be abandoned, then we might wonder if there is a radical change to our methods that would allow us to measure it.<sup>26</sup> Although I can only gesture at the reasons here, I suspect such a revolution will never come. I think Virtue Ethicists like Rosalind Hursthouse (1999, p. 56) are correct that wise understanding is 'uncodifiable': we cannot boil down the decisions that comprise a well-lived life to a set of rules that an unwise person could use to decide what they ought to do in all the situations they might face. If that's the case, it will be challenging to develop clear and representative success conditions for particular all-things-considered decisions. This provides a marked contrast between practical wisdom and other domains where we have successfully operationalized expertise and its components. The vast variety of board arrangements and moves in chess do not impede measurement precisely because the end – a winning board position – *is* codifiable.

So, though I am conscious of all the cases where skeptics have been proven wrong in the past, I suspect that practical wisdom is not the sort of thing we should expect to measure.

#### 5.4 In support of a change of focus

If practical wisdom is a worthy target but we can't measure it, what does that leave? Importantly, giving up on measuring wisdom does not mean we have to give up on studying general features of practical wisdom. An important analogue in the history of science will illustrate.

A perplexing thing about Darwinian evolutionary theory is that it does an impressive job of explaining the development of organisms, but it doesn't help us predict which organisms will actually develop or what they'll be like (Scriven, 1959). We can use it to explain why a strange creature like the platypus arose, but it will not help us predict with any confidence which specific creatures will exist in uncontrolled, real-world places in the future. Put simply, evolutionary theory does well with explanation but not well with prediction (1959, p. 131). This does not speak against the value or plausibility of the theory, but it does indicate that, like most tools, it is valuable for some projects but not others.

Similarly, saying that we can't measure practical wisdom doesn't entail the demise of interdisciplinary research on practical wisdom. The study of practical wisdom shares important features with the study of evolutionary processes. Of course, there are differences. Unlike evolutionary theories, an account of practical wisdom does not aim merely to describe how things *are*; instead, practical wisdom is a prescriptive ideal that tells us how we *ought* to live, and this ideal cannot be described with empirical science alone (Swartwood & Tiberius, 2019, pp. 14–20). But wisdom science and evolutionary theory both study objects of great complexity, in which the variables are so many and interact in such complex ways that modeling them in any detail is extremely challenging, if not practically impossible. The complexity of factors that contribute to how organisms evolve or change are matched by the complexity of factors that contribute to a decision, life, or person exhibiting practical wisdom. So, if evolutionary theory's inability to predict does not warrant sorrow or surprise, neither should our inability to operationalize practical wisdom.

There are a number of interdisciplinary methods for examining general features of practical wisdom, even if we cannot measure it (Swartwood & Tiberius, 2019, pp. 20–34). In some cases, changes made will be less to the methods of philosophy or social science than to how we conceive of and combine their results. Philosophical reasoning can help us specify which practical wisdom-relevant characteristics (e.g. reasoning or discrimination skills or processes, self-regulation skills or habits) are logically entailed by a rationally defensible account (such as the minimal philosophical conception). Empirical research can then determine how these can be developed and how they relate to other characteristics whose necessity for practical wisdom is more contentious or contingent on facts about human psychology. (For example, does an ability to give advice or justify one's choices tend to develop as a result of acquiring the reasoning and self-regulation processes that are part of a philosophically plausible picture of wisdom?) Even if the results of such an investigation would not be sufficient to allow us to measure practical wisdom or its components, they still would provide valuable information about some of the necessary conditions for having or developing practical wisdom and a general picture of some of the ways it would be likely to manifest in real human lives. (Compare general best practice measures of teaching: they can help us acquire a fuller picture of good teaching and how we can improve, even if they do not constitute measures of teaching expertise and should not be used as such.)

## 6. CONCLUSION

This is an exciting time for those interested in the interdisciplinary study of wisdom. Combining the tools of philosophy and science promises a deeper and more plausible picture of how wisdom would manifest and could be cultivated in real people. But making good on this promise requires acknowledging and addressing disciplinary differences in how wisdom is conceptualized. The argument defended here lays bare one such difference: whatever contemporary measures are operationalizing, it is not a philosophically plausible practical wisdom or its components. My hope is that examining this argument and its implications will produce a more unified and focused vision for the interdisciplinary study of wisdom.<sup>27</sup>

### **References:**

- Annas, J. (2004). Being virtuous and doing the right thing. *Proceedings and Addresses of the American Philosophical Association*, 78, 61–75.
- Annas, J. (2011). *Intelligent virtue*. Oxford, UK: Oxford University Press.
- Ardelt, M. (2003). Empirical assessment of a three-dimensional wisdom scale. *Research on Aging*, 25(3), 275–324.
- Ardelt, M. (2004). Wisdom as expert knowledge system: A critical review of a contemporary operationalization of an ancient concept. *Human Development*, 47(5), 257–285.
- Aristotle. (1999). *Nicomachean Ethics* (T. Irwin, Trans.). Hackett Publishing Company.
- Baltes, P. B., & Staudinger, U. M. (2000). Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. *American Psychologist*, 55(1), 122.
- Bassett, C. L. (2011). Understanding and teaching practical wisdom. *New Directions for Adult and Continuing Education*, 2011(131), 35–44.
- Berger, P., & Luckmann, T. (1991). *The social construction of reality*. London, England: Penguin Books.
- Bluck, S., & Glück, J. (2005). People's Implicit Theories of Wisdom. *A Handbook of Wisdom: Psychological Perspectives*, 84.
- Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*. Retrieved from <https://www.scienceopen.com/document?id=25ff22be-8a1b-4c97-9d88-084c8d98187a>
- Broadie, S. (1993). *Ethics with Aristotle*. New York, NY: Oxford University Press.
- Brooks, J., & Walsh, P. (2017, May 11). St. Olaf: Report of racist note on black student's windshield was "fabricated." Retrieved July 3, 2019, from StarTribune website: <http://www.startribune.com/st-olaf-report-of-racist-note-on-black-student-s-windshield-was-fabricated/421912763/>
- Darnell, C., Gulliford, L., Kristjánsson, K., & Paris, P. (2019). Phronesis and the Knowledge-Action Gap in Moral Psychology and Moral Education: A New Synthesis? *Human Development*, 1–29.
- Elo, A. E. (1986). *The rating of chessplayers, past and present* (2nd ed.). New York, NY: Arco.
- Ericsson, K. A. (2018). An Introduction to the Second Edition of The Cambridge Handbook of Expertise and Expert Performance: Its Development, Organization, and Content. In *Cambridge Handbook of Expertise and Expert Performance* (2nd ed.). Retrieved from [http://sptproxy.mnpals.net/login?url=https://search.credoreference.com/content/entry/cupexpert/an\\_introduction\\_to\\_the\\_secon](http://sptproxy.mnpals.net/login?url=https://search.credoreference.com/content/entry/cupexpert/an_introduction_to_the_secon)

d\_edition\_of\_the\_cambridge\_handbook\_of\_expertise\_and\_expert\_performance\_its\_development\_organization\_and\_content/0?institutionId=5403

- Foot, P. (2003). *Natural goodness*. Clarendon Press.
- Glück, J. (2017). Measuring wisdom: Existing approaches, continuing challenges, and new developments. *The Journals of Gerontology: Series B*, 73(8), 1393–1403.
- Gobet, F., & Charness, N. (2018). Expertise in Chess. In *Cambridge handbook of Expertise and Expert Performance* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52(11), 1209.
- Grimm, S. R. (2015). Wisdom. *Australasian Journal of Philosophy*, 93(1), 139–154.
- Grossmann, I. (2017). Wisdom in context. *Perspectives on Psychological Science*, 12(2), 233–257.
- Grossmann, I., Dorfman, A., & Oakes, H. (2019). Wisdom is a social-ecological rather than person-centric phenomenon. *Current Opinion in Psychology*.
- Grossmann, I., Kung, F. Y., & Santos, H. C. (2019). Wisdom as state versus trait. In *The Cambridge Handbook of Wisdom* (pp. 249–274). Cambridge, UK: Cambridge University Press.
- Grossmann, I., Na, J., Varnum, M. E., Park, D. C., Kitayama, S., & Nisbett, R. E. (2010). Reasoning about social conflicts improves into old age. *Proceedings of the National Academy of Sciences*, 107(16), 7246–7250.
- Gump, S. E. (2007). Student Evaluations of Teaching Effectiveness and the Leniency Hypothesis: A Literature Review. *Educational Research Quarterly*, 30(3), 56–69.
- Haybron, D. (2007). Life satisfaction, ethical reflection, and the science of happiness. *Journal of Happiness Studies*, 8(1), 99–138.
- Haybron, D. M. (2008). *The pursuit of unhappiness: The elusive psychology of well-being*. New York, NY: Oxford University Press, USA.
- Haybron, D. M. (2011). Taking the satisfaction (and the life) out of life satisfaction. *Philosophical Explorations*, 14(3), 249–262.
- Hessler, M., Pöpping, D. M., Hollstein, H., Ohlenburg, H., Arnemann, P. H., Massoth, C., ... Wenk, M. (2018). Availability of cookies during an academic course session affects evaluation of teaching. *Medical Education*, 52(10), 1064–1072.  
<https://doi.org/10.1111/medu.13627>
- Hursthouse, R. (1999). *On virtue ethics*. New York, NY: Oxford University Press.
- Hursthouse, R., & Pettigrove, G. (2016). Virtue Ethics. In *Stanford Encyclopedia of Philosophy* (Winter 2016). Retrieved from <https://plato.stanford.edu/entries/ethics-virtue/>
- Ivanhoe, P. J. (2002). Confucian Self Cultivation and Mengzi's Notion of Extension. In X. Liu & P. J. Ivanhoe (Eds.), *Essays on the Moral Philosophy of Mengzi* (pp. 221–241).
- Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. *Research Paper. MET Project. Bill & Melinda Gates Foundation*. Retrieved from <https://files.eric.ed.gov/fulltext/ED540959.pdf>

- Kraut, R. (1993). In Defense of the Grand End. *Ethics*, 103(2), 361–374.
- Kraut, R. (2009). *What is good and why: The ethics of well-being*. Cambridge, MA: Harvard University Press.
- Kristjánsson, K. (2015). *Aristotelian character education*. Routledge.
- Kunzmann, U. (2019). Performance-based measures of wisdom: State of the art and future directions. In *The Cambridge Handbook of Wisdom* (pp. 277–296). Cambridge, UK: Cambridge University Press.
- Kunzmann, Ute, & Baltes, P. B. (2003). Wisdom-related knowledge: Affective, motivational, and interpersonal correlates. *Personality and Social Psychology Bulletin*, 29(9), 1104–1119.
- McDowell, J. (1979). Virtue and reason. *Monist*, 62(3), 331–350.
- McDowell, J. (1996). Deliberation and moral development in Aristotle's ethics. In *Aristotle, Kant and the Stoics: Rethinking Happiness and Duty* (pp. 19–35). New York, NY: Cambridge University Press.
- Mengzi. (2008). *Mengzi: With selections from traditional commentaries* (B. W. Van Norden, Trans.). Indianapolis, IN: Hackett Publishing Company.
- Moxley, J. H., & Charness, N. (2013). Meta-analysis of age and skill effects on recalling chess positions and selecting the best move. *Psychonomic Bulletin & Review*, 20(5), 1017–1022. <https://doi.org/10.3758/s13423-013-0420-5>
- Nussbaum, M. C. (2001). *The fragility of goodness: Luck and ethics in Greek tragedy and philosophy*. Cambridge, UK: Cambridge University Press.
- Nussbaum, M. C. (2008). Non-relative virtues: An Aristotelian approach. *Midwest Studies in Philosophy*, 13(1), 32–53.
- Oakes, H., Brienza, J., Elnakouri, A., & Grossmann, I. (2018). *Wise reasoning: Converging evidence for the psychology of sound judgment*.
- Richardson, H. S. (1990). Specifying norms as a way to resolve concrete ethical problems. *Philosophy & Public Affairs*, 19(4), 279–310.
- Richardson, H. S. (2000). Specifying, balancing, and interpreting bioethical principles. *Journal of Medicine and Philosophy*, 25(3), 285.
- Scriven, M. (1959). Explanation and prediction in evolutionary theory. *Science*, 130(3374), 477–482.
- Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, 136(2), 253–263.
- Sternberg, R. J. (1998). A balance theory of wisdom. *Review of General Psychology*, 2(4), 347.
- Sternberg, R. J., & Glück, J. (2018). *The Cambridge handbook of wisdom*. Cambridge University Press Cambridge.
- Stichter, M. (2016). The Role of Motivation and Wisdom in Virtues as Skills. In *Developing the Virtues: Integrating Perspectives*. New York, NY: Oxford University Press.
- Stigler, J. W., & Miller, K. F. (2018). Expertise and Expert Performance in Teaching. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & M. A. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 431–452). Retrieved from [http://sptproxy.mnpals.net/login?url=https://search.credoreference.com/content/entry/cupexpert/expertise\\_and\\_expert\\_performance\\_in\\_teaching/0?institutionId=5403](http://sptproxy.mnpals.net/login?url=https://search.credoreference.com/content/entry/cupexpert/expertise_and_expert_performance_in_teaching/0?institutionId=5403)
- Swartwood, J. D. (2013). Wisdom as an Expert Skill. *Ethical Theory and Moral Practice*, 16(3), 511–528. <https://doi.org/10.1007/s10677-012-9367-2>

- Swartwood, J. D., & Tiberius, V. (2019). Philosophical Foundations of Wisdom. In R. J. Sternberg & J. Glück (Eds.), *The Cambridge Handbook of Wisdom* (pp. 10–39). Cambridge, UK: Cambridge University Press.
- Thomas, S., & Kunzmann, U. (2013). Age differences in wisdom-related knowledge: Does the age relevance of the task matter? *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 69(6), 897–905.
- Tiberius, V. (2003). *How's It Going?" Judgments of Overall Life-Satisfaction and Philosophical Theories of Well-Being*.
- Tiberius, V. (2006). Well-being: Psychological research for philosophers. *Philosophy Compass*, 1(5), 493–505.
- Tiberius, V. (2008). *The reflective life: Living wisely with our limits*. New York, NY: Oxford University Press.
- Tiberius, Valerie. (2004). Cultural differences and philosophical accounts of well-being. *Journal of Happiness Studies*, 5(3), 293–314.
- Tiberius, Valerie. (2013). Philosophical methods in happiness research. In *Oxford handbook of happiness*.
- Webster, J. D. (2019). Self-report wisdom measures: Strengths, limitations, and future directions. In *The Cambridge Handbook of Wisdom* (pp. 297–320). Cambridge, UK: Cambridge University Press.
- Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors*, 45(1), 104–116.
- Weiss, D. J., & Shanteau, J. (2014). Who's the best? A relativistic view of expertise. *Applied Cognitive Psychology*, 28(4), 447–457.
- Weststrate, N. M., Bluck, S., & Glück, J. (2019). Wisdom of the crowd: Exploring people's conceptions of wisdom. *The Cambridge Handbook of Wisdom*, 97–121.
- Wolf, S. (2007). Moral Psychology and the Unity of the Virtues. *Ratio*, XX(2), 145–167.

<sup>1</sup> For a comprehensive overview of some of the main lines of research, see Sternberg and Glück (2018).

<sup>2</sup> For instance, Grossmann et al (2019) explicitly claim to “adopt the notion of practical wisdom (cf. *phronesis*; Aristotle) – a form of excellence in ethical and practical deliberation about the best course of action in a complex social situation.”

<sup>3</sup> I will use ‘they’ for the third person singular to acknowledge those who prefer neither ‘he’ nor ‘she.’

<sup>4</sup> Weiss and Shanteau (Weiss & Shanteau, 2014, p. 3) helpfully illustrate the importance of taking base rate considerations into account: you wouldn't be measuring expertise if you only test a chess player's rate of winning against highly inexperienced opponents or if you measure a weather forecaster's ability to predict an obvious and easily observable weather pattern (such as a tendency not to get rain in a dry desert area). A similar point applies to the all-things-considered decisions that are the domain of wisdom.

<sup>5</sup> Unlike chess expertise, expert teaching cannot be demonstrated without at least some minimal ability and motivation on the part of others – in this case, the student (Stigler & Miller, 2018, p. 431).

<sup>6</sup> This is thus an instance of the general problems with identifying expertise through reputation and “social acclamation” (Ericsson, 2018; Shanteau, Weiss, Thomas, & Pounds, 2002, p. 255).

<sup>7</sup> Some might worry this claim implies we can't measure things like intelligence, given the variety of different definitions of different types of intelligence that have been offered. But this misses the point. The claim here does not imply that we can't measure intelligence, for the same reasons that disagreements about the goals of teaching do not necessarily imply we can't measure teaching expertise. My point is not that a disagreement about how to define a construct undermines measuring it. Instead, the point is we can be confident we're measuring a particular definition of a construct only if the measure is specified by clear and specific success conditions consistent with that definition. We can do that with various different definitions of intelligence, even if we'll then wonder which definitions really matter to us. I thank Judith Glück for pointing me to this possible objection.

<sup>8</sup> I am using ‘ought’ and ‘should’ as synonyms here.

<sup>9</sup> Some might resist this point and insist that “wisdom is ... in the eye of the beholder” (Glück, 2017, p. 1400) or that wisdom is a “cultural construct” (Bluck & Glück, 2005, pp. 90–91). This may be motivated in part by prominent views of social construction utilized in sociology or psychology (Berger & Luckmann, 1991). If that's true, then perhaps research on implicit theories of wisdom (lay or folk views of wisdom) could shed some light on what counts as practical wisdom and what doesn't.

There are at least two reasons that this strategy will not help us specify the success conditions for the philosophically plausible account of practical wisdom that I am outlining. First, implicit theories of wisdom tell us what (or whom) people *consider* wise, but that's not a good guide to what's *actually* wise (Swartwood & Tiberius, 2019). Second, research on implicit theories does not usually give much information about what people think ought to be done, all-things-considered, in particular cases. I thank Andrew Kubas for helpful discussion of these points.

<sup>10</sup> I'm not intending this point to be a highly theoretical one. But, for examples of theoretical elaborations on these themes, see Broadie (1993), Kristjánsson (2015), Richardson (1990, 2000), Stichter (2016), and Tiberius (2008).

<sup>11</sup> Some of the decisions will also be about which situations to put oneself into and which to avoid.



---

<sup>12</sup> Compare a real life case, described in Brooks & Walsh (2017).

<sup>13</sup> For instance, psychologist Caroline Bassett (2011, p. 304) describes wise judgment as “a kind of excellence in judgment where cognitive, affective, reflective, and active qualities all work together in harmony to produce decisions or behaviors that lead to a common, general social good.” I’m not claiming that Bassett’s definition is representative of wisdom scientists; the point is merely to show that the temptation I’m describing is a real one.

<sup>14</sup> Questions about what’s good for a particular person – what makes their life go well for them – concern well-being. For discussion of the variety of competing philosophical accounts of well-being and how they relate to the study of well-being and related concepts in psychology, see, for example, Haybron (2007; 2008, 2011) and Tiberius (2003, 2006; 2004, 2013).

<sup>15</sup> I am not denying that we can identify plausible *prima facie* or *pro tanto* moral principles. But, I agree with Hursthouse (1999, pp. 56–57) that even a plausible and complete set of principles would require significant wisdom to apply to particular situations.

<sup>16</sup> Darnell et al. state that “phronesis properly yields decisions ..., each of which embodies a correct prescription or right reason for a given set of circumstances, which are context-sensitive, that is, they vary with the features of the situation and the individual experiencing the situation” (Darnell et al 2019, p. 13). This acknowledges that wisdom deals with how we ought to live and conduct ourselves. And, the focus on particular circumstances seems to acknowledge that whether you have practical wisdom depends on whether you can decide well what to do in particular circumstances. In addition, they also emphasize that “the phronimos’ deliberation aims at the unqualified good, or what is good – all things considered” (p. 13).

<sup>17</sup> Stichter’s view is part of a response to the puzzle of the unity of the virtues. For an explanation of the puzzle (and some of the prominent responses) intended to be accessible to non-philosophers, see Swartwood and Tiberius (2019, pp. 23–25).

<sup>18</sup> Other issues the conception is neutral on include: whether wisdom comes in degrees or not, the degree of particularism governing wise decisions, in what sense there is a unity of the virtues, whether there are any situations posing irresolvable moral dilemmas, whether we can measure absolute or merely relative levels of wisdom, whether a practically wise person will have to use self-regulation skills to overcome temptations to do the wrong thing, and whether there are multiple equally good decisions in some particular situations.

<sup>19</sup> Compare Stigler and Miller (2018): “Because teaching is highly contextual and complex, teachers must also have the ability to decide which of many possible strategies they should pursue in any given time and place. They need to be able to size up a situation, decide which strategy to employ, and then adapt it to achieve their specific instructional goal. In other words, teachers need judgment. They can be highly knowledgeable in all the ways Shulman (1987) describes, and highly skilled at implementing a wide variety of instructional strategies. But unless they make good decisions about when and how to employ their knowledge and skill, their knowledge may not serve to support students’ learning.”

<sup>20</sup> I would like to thank Monika Ardelt for helpful discussion that pushed me to clarify my point here.

<sup>21</sup> This is why moral philosophers, following Aristotle, often describe the virtue of compassion (as opposed to the personal psychological characteristic) as the disposition to respond to threats to others’ well-being *at the right times, in the right ways, and for the right reasons* (Aristotle, 1999, I. 1109a25; Hursthouse & Pettigrove, 2016). Note that a virtue in this sense could be defined as a trait, a state, or a combination of the two (though I think the last two options are most plausible).

<sup>22</sup> Cases like these led Aristotle to distinguish between practical wisdom and mere cleverness. While the merely clever person has great skill at determining how to achieve the goals they happen to have, a person with practical wisdom has the same skill directed at the right goals, at the right times, and for the right reasons (Aristotle, 1999, II. 1144a25-36).

<sup>23</sup> For an overview of other measures, see Kunzmann (2019), Webster (2019), and Glück (2017). For an overview of an interesting interdisciplinary model of practical wisdom for which a measure is currently under development, see Darnell et al. (2019). At the time of this writing, I am not aware of any measures that avoid the problems I have identified for the measures focused on in section 2 of this paper.

<sup>24</sup> Some might argue that premise 1 of the argument is false, because it ignores alternative methods for measuring expertise when success conditions for the domain cannot be adequately identified. For instance, a prominent and well-researched example of a measure of expertise that doesn’t require success conditions is the CWS (Cohran Weiss Shanteau) ratio (Shanteau et al., 2002; Weiss & Shanteau, 2003, 2014). The CWS ratio is based upon the idea that expert judgment must be both discriminating and consistent (Shanteau et al., 2002, p. 258). An expert must be discriminating in the sense that they make relevant distinctions between different situations when deciding what to do. An expert must also be consistent in the sense that they make similar judgments in similar situations. The CWS ratio is supposed to capture these insights and allow us to measure expertise even in cases where we lack a ‘gold standard’ of success with which to compare a putative expert’s performance (Shanteau et al., 2002, p. 253). The higher the ratio of discrimination to inconsistency, the more the person acts like an expert, relative to other performers in the same situations (Shanteau et al., 2002, p. 258; Weiss & Shanteau, 2014, p. 4). Measures that utilize the CWS ratio can thus measure expertise even when we do not have clear conditions for successful performance.

This objection is not persuasive. The argument I’ve defended relies upon the idea that expert performance is a necessary condition of expertise. This idea is both intuitively compelling and has been behind much of the influential empirical research on expertise. The CWS, while promising, does not make a clear account of success unnecessary for measuring expertise. Its proponents admit that demonstration of high discrimination and low inconsistency may be necessary for expertise, but it is not sufficient (Weiss & Shanteau, 2014, p. 3). Someone may very consistently apply the same general rules to all their chess moves, and these may allow them to make a variety of subtle discriminations. But, they still may end up reliably losing their games. In psychological terms, applications of the CWS do not guarantee validity – they do not guarantee that what is being measured (the CWS ratio) is what is actually supposed to be measured (genuine expertise) (Weiss & Shanteau, 2014, p. 3).

Indeed, I think the situation is even worse when we consider the domain of wisdom. Part of what is challenging in all-things-considered situations is determining which features of a situation are relevant and which are not. Determining that would be necessary even before we could determine which decisions are similar in the relevant ways and which are not. (Which decision in Benevolent Scapegoating would be relevantly similar to which decision in Caretaker Fib or Undiagnosed?) That philosophical work of determining which features are relevant needs to be done before the CWS could even be applied. I suspect other methods for measuring high-level performance in the absence of clear success conditions would have the same problems when applied to practical wisdom.

<sup>25</sup> I thank Monika Ardelt and Matt Stichter for pushing me to clarify this point.

<sup>26</sup> I am grateful to Judith Glück, Stephen Grimm, and Matt Stichter for helpful discussion of this point.

<sup>27</sup> To the extent that I have succeeded in making this argument accessible and compelling to an interdisciplinary audience, I owe thanks to many people who have graciously lent me their feedback and expertise. The audience at the Explaining Wisdom conference, in Dresden on 28 June 2019, provided much helpful feedback on and discussion of an initial version of the argument. Judith Glück, Andrew Kubas, Ian Stoner, and Ruth

---

Swartwood provided insightful feedback on various drafts of the paper. Ian Stoner provided, on several occasions, especially helpful feedback on the organization of the paper and details of the presentation of the argument. Valerie Tiberius provided much needed encouragement at critical points. Three reviewers provided generous feedback that helped me make the argument more precise and more accessible to an interdisciplinary audience.