Phronesis

Kristjansson, Kristjan; McLoughlin, Shane; Thoma, Steve

UNIVERSITY OF BIRMINGHAM | THE JUBILEE CENTRE FOR CHARACTER & VIRTUES

PHRONESIS:

DEVELOPING AND VALIDATING A SHORT MEASURE OF PRACTICAL WISDOM

RESEARCH REPORT

KRISTJÁN KRISTJÁNSSON
SHANE McLOUGHLIN
STEPHEN THOMA

www.jubileecentre.ac.uk

Jubilee Centre for Character and Virtues

The Jubilee Centre for Character and Virtues is a unique and leading centre for the examination of how character and virtues impact on individuals and society. The Centre was founded in 2012 by Professor James Arthur. Based at the University of Birmingham, it has a dedicated team of over 20 academics from a range of disciplines, including: philosophy, psychology, education, theology and sociology.

With its focus on excellence, the Centre has a robust, rigorous research and evidence-based approach that is objective and non-political. It offers world-class research on the importance of developing good character and virtues and the benefits they bring to individuals and society. In undertaking its own innovative research, the Centre also seeks to partner with leading academics from other universities around the world and to develop strong strategic partnerships.

A key conviction underlying the existence of the Centre is that the virtues that make up good character can be learnt and taught. We believe these have been largely neglected in schools and in the professions. It is also a key conviction that the more people exhibit good character and virtues, the healthier our society. As such, the Centre undertakes development projects seeking to promote the practical applications of its research evidence.

> TO GRASP WHAT *PHRONESIS* IS, WE SHOULD FIRST STUDY THE SORT OF PEOPLE WE CALL *PHRONETIC*
>
> *Aristotle, Nicomachean Ethics, 1140a25-26*

University of Birmingham 2023
ISBN: 9780704429857

PHRONESIS:

DEVELOPING AND VALIDATING A SHORT MEASURE OF PRACTICAL WISDOM

RESEARCH REPORT

Contents

Executive Summary

Phronesis (practical wisdom) has come under increased scrutiny of late within neo-Aristotelian moral psychology, character education and virtue-based professional ethics. While great strides have been made within the Jubilee Centre since 2019 in understanding the concept of *phronesis* and creating a theoretically viable instrument to measure it, the instrument originally designed by the Centre, the Long *Phronesis* Measure (LPM), turned out to be practically unwieldy, in the sense of taking too long to complete (over 45 minutes) and being complicated to score.

The aim of the final phase of the 2019–2023 *Phronesis* Project, reported upon here, was to go back to the drawing board and create an easier-to-use instrument, referred to as a Short *Phronesis* Measure (SPM), while retaining as much as possible the theoretical considerations that motivated the original measure.

This report:

- Explores the proposed conceptual contours of *phronesis*, and charts the journey from the Jubilee Centre's four-componential Aristotelian *Phronesis* Model (APM) to the Long *Phronesis* Measure (LPM), created in the early stages of the *Phronesis* Project, with a special focus on its practical shortcomings.
- Describes four empirical studies with large UK (N = 2000; N = 1000) and US (N = 1000) samples that helped create a viable Short *Phronesis* Measure (SPM).
- Discusses and contextualises the new measure.
- Paves the way for further practical research and suggests the next steps in further strengthening the measure.

Key findings:

- Exploratory and confirmatory factor analyses helped create a theoretically credible and practically viable measure of *phronesis*, which should not take more than 20 minutes to complete.
- The data confirm acceptable psychometric properties overall: a good fit with a three-component model of *phronesis* (including the emotion-regulation, blueprint/moral identity and adjudication components), and also an adequate fit with a four-component model (including the constitutive/moral perception component); hence supporting in all essentials the original Jubilee Centre's APM conceptualisation.
- The three factors in the best-fit three-factor model predict flourishing, as Aristotle would have anticipated, apart from that of financial security. This adds considerable backbone to the APM, especially because flourishing was not explored as part of the development of the original LPM.
- While the findings from the UK and US samples mostly coincide in terms of model fit, US participants reported higher levels of *phronesis* across almost all its variables. The reasons for this apparent cultural difference are not entirely clear at present.
- Various subsidiary measures were administered to understand more fully the workings of *phronesis* and its relationships to other personality and characterological variables; but those await further analysis and are beyond the scope of the current report.

The large data set gathered in the summer of 2023 has not yet been fully analysed, especially with respect to findings from the various subsidiary measures administered. Once that work has been completed, and results subjected to academic peer-review, the new SPM will be made available for free on the Jubilee Centre website, along with a scoring key. Meanwhile, those who might be interested in making use of the SPM for exploratory purposes before its formal launch can contact Dr. Shane McLoughlin (s.mcloughlin@bham.ac.uk).

1 Purpose of the Report

This is the third and final research report emanating from the *Phronesis* Project, supported by the John Templeton Foundation and run in the Jubilee Centre between 2019 and 2023.*[1]* The aim of the project has been to explore the factors that motivate moral action and, hence, bridge the gap between knowing the good and doing the good. In particular, the project has aimed at exploring the role that Aristotle's intellectual meta-virtue of *phronesis* or practical wisdom plays in moral adjudication and decision-making. Despite a recent surge of interest in *phronesis* predating this project, a mismatch remained between the interest in the construct and any serious attempts to specify and evaluate it. Thus, no rigorous measurable conceptualisation of *phronesis* existed prior to 2019 and no psychological instrument had yet been designed to measure it.

Despite the relative success of the Jubilee Centre in piloting an apparently viable measure of *phronesis* (Kristjánsson *et al.*, 2020; Darnell *et al.*, 2022), various practical problems remained that motivated an extensive revision and conceptual replication of previous research work, aimed at developing a shorter and more practicable measure. Those problems are explained in Section 2.5. In what follows, the original measure from 2020 will be referred to as the Long *Phronesis* Measure (LPM) and the measure refined and validated in this final phase of the project as the Short *Phronesis* Measure (SPM).

In addition to the practical problems encountered in implementing LPM, some troublesome theoretical issues persisted. While the original LPM was tested only on small convenience samples of U.K. secondary-school students and young adults, among the issues still unaddressed was, first, whether such a *phronesis* measure would also have traction in (a) larger samples, including (b) participants outside of the U.K. and (c) nationally representative adult samples. Second, the earlier piloting took the JC conceptualisation of *phronesis* (see Section 2.3) for granted and began the study with a confirmatory factor analysis (i.e., an analysis confirming the hypothesised components of Aristotelian *phronesis)*. Some academics are wary of such a top-down approach to instrument design, especially in the case of a newly conceptualised construct, and would recommend a bottom-up approach, beginning with an exploratory factor analysis, i.e. an analysis in which the 'mesh' shapes and sizes of the 'net' used to catch the construct are not pre-determined by theory. Third, for reasons of time and convenience, we decided in the earlier study to lift, as much as possible, measures of the putative components of *phronesis* from already validated off-the-shelf instruments. While this saved time, in terms of not having to validate the pre-existing measures, some compromises had to be made. For example, the measures used to identify the emotion-regulative component of *phronesis* focused mostly on tracking empathy, which is arguably a precursor of appropriate emotions but does not specifically target the regulative element. Fourth, our earlier study used self-reported prosocial behaviour as an outcome variable. Although it does of course matter that the *phronesis* construct was found to predict this outcome variable – especially in the context of addressing the infamous 'gappiness problem' in moral psychology (see Section 2.4) – it would be more salient, from an Aristotelian perspective (as Aristotle had no concept of prosociality), to explore whether *phronesis* is associated with greater overall flourishing (*eudaimonia*). Fifth, the earlier study controlled for the Big-Five personality traits, but it did not study the relationship between *phronesis* and various other potentially relevant traits, such as the Dark Tetrad, Honesty-Humility, and measures of Moral Foundations.[2]

In light of those remaining issues, the research questions that guided the current research work were:

- Does a new exercise in the design of a shorter *phronesis* instrument, using a greater number of participants of varied ages inside and outside of the U.K., and starting with a bottom-up exploratory factor analysis, confirm the putative viability of the Jubilee Centre's *phronesis* construct?
- Does a hierarchical confirmatory factor analysis, in which the factors extracted from the exploratory factor analysis are considered sub-factors of the conceptualised components of *phronesis*, fit the data?
- Does the model extracted from the confirmatory factor analysis predict a latent flourishing variable?
- More generally, can the Jubilee Centre produce an instrument to measure *phronesis* that is shorter and easier than the earlier incarnation?

The current report does not purport to offer definitive answers to all these questions, nor to possible subsidiary research questions that relate to other potential correlations in this large data set. However, we consider this report to respond powerfully to worries raised about the shortcomings of the earlier LPM for practical purposes.

[1] https://www.jubileecentre.ac.uk/1756/projects/phronesis
[2] Findings from the various subsidiary measures utilised in the present study are not reported upon in the current report, which focuses on the development of the SPM as a practicable instrument.

2 Background

2.1 THE *PHRONESIS* BANDWAGON

Neo-Aristotelian character or virtue education has been undergoing a revival of late, either as a form of values/moral education (Jubilee Centre, 2022) or part of citizenship education focused on the development of civic virtues (Peterson, 2020). In part motivated by a new international policy-drive towards seeing flourishing (*eudaimonia* in Aristotle's sense) as the ultimate aim of education (Kristjánsson, 2020), this new-found interest has inter alia led to the establishment of a European Character and Virtue Association.

In Aristotle's ethical and educational system, the lynchpin of a flourishing life, actualising the virtues and representing good character, is the overarching meta-virtue of *phronesis* (practical wisdom): an intellectual virtue that guides the moral and civic virtues towards their goals and solves possible conflicts between them as an integrator and adjudicator. According to Aristotelian character developmental theory, young people who have acquired the right moral traits through habituation and role modelling need gradually to develop this intellectual virtue to guide their decision-making. Otherwise, their moral lives will be fragmented, uncritical and lacking in intrinsic value. In that sense, then, *phronesis* is best understood as excellence in ethical decision-making (Kristjánsson and Fowers, 2024).

It is somewhat mysterious, however, that until recently much less was written about *phronesis* as a meta-virtue than about the underlying primary virtues in philosophical and educational circles, and prior to work in the Jubilee Centre in 2019 no psychologically credible conceptualisation of *phronesis* existed, nor any instrument to measure its efficacy (Darnell *et al.*, 2019). This lacuna is particularly striking within education where advice about how to cultivate *phronesis*, in schools or universities, has been in short supply (Kristjánsson, 2021). Explanations given for this academic void range from Aristotle's own reticence about *phronesis* cultivation to the fact that *phronesis* is a more complex construct than, say, a 'simple' moral virtue like gratitude.

In any case, the last 3-4 years have witnessed a sudden burst of interest in *phronesis* and *phronesis* development within philosophy, psychology, professional ethics and education, with a number of partly overlapping constructs of *phronesis* being created (De Caro *et al.*, 2021; Wright *et al.*, 2021; Fowers *et al.*, 2021; Kristjánsson *et al.*, 2021; Darnell *et al.*, 2022). There is no space here for comparisons and contrasts between all the different constructs. Rather, we aim to retrace the steps that led to the creation of the streamlined SPM instrument described in the present report. That aim prompts a rehearsal, below, of the so-called Aristotelian 'standard model' of *phronesis*, followed by the more detailed conceptualisation of an Aristotelian *phronesis* model created by the Jubilee Centre. We summarise some of the main results of the previous two reports emanating from this project and end with the overall evaluative goals that drove this final phase of the *Phronesis* Project.

2.2 THE ARISTOTELIAN STANDARD MODEL

Phronesis is, as already noted, a key concept in Aristotelian and Aristotle-inspired theories of moral and character education. 'Character education' here refers to the cultivation of positive individual traits that are conducive to and constitutive of human flourishing, individual and societal (Jubilee Centre, 2022), and Aristotelians call those traits 'virtues'. In short, *phronesis* refers to the capacity of knowing and enacting the right course of (moral) action through a process of identifying and deliberating between competing values, emotions and alternatives. It:

- is a virtue of autonomous, critical thinking;
- deals with human action;
- consists of both instrumental cleverness and already habituated virtues;
- involves excellence in practical deliberation.

Aristotle's idea is that we all possess different sets of virtues – moral, civic, intellectual and performative (Jubilee Centre, 2022). However, the demands of these virtues often come into conflict with one another, between sets or within sets. For example, it is difficult enough to learn how to be honest. It is even more difficult, however, to know what to do when honesty clashes with considerateness. It is then that we need *phronesis* for arbitration.

There are already large theoretical literatures on practical wisdom in general (e.g., Russell, 2009) and the Aristotelian concept in particular (e.g., Curzer, 2012). However, most of those literatures are either exegetical or purely philosophical in orientation, and hence outside of our immediate practical interests. What matters for present purposes is that in philosophy there has gradually evolved what Miller (2021) calls a neo-Aristotelian 'standard model of *phronesis'*, which carries independent interest, whatever one may think of some of Aristotle's own claims. This model owes a lot to Russell's (2009) meticulous analysis of the concept of practical wisdom, although the model he excavates is not meant to be exclusively Aristotelian.

In the standard neo-Aristotelian model, the task of *phronesis* is complex (Tiberius and Swartwood, 2011), and a common suggestion from the literature is that it has at least three components. First, in the constitutive function of single-virtue-application, *phronesis* helps the (budding) *phronimos* to spot situations where the relevant virtue is required and how to execute it. For example, courage is the virtue that is appropriate to situations involving risk. Second, the integrative function of conflicting-virtues arbitration allows the *phronimos* to integrate different virtues that seem to come into conflict in the same situation, such as being courageously generous. This arbitration can also lead to enacting one virtue that is a higher priority and in unresolvable conflict with a second virtue (e.g., mercy versus justice). Only through *phronesis* do the virtues become a 'package deal' (Russell, 2009, p. 26). Third, the function of emotion regulation builds on emotional dispositions cultivated through habituation, in that the *phronimos* re-evaluates those early dispositions critically, infusing them with reason and justification. Others have added a fourth, the function of 'deep understanding' (Tiberius and Swartwood,

2011) of the human condition, to this mix: an understanding of what constitutes human flourishing as an irreducibly moral activity. We have followed that lead and refined it in our detailed four-componential model, explained in the following sub-section.

2.3 UNPACKING THE STANDARD MODEL: A CONCEPTUALISATION OF ARISTOTELIAN *PHRONESIS*

Most of Aristotle's discussion of *phronesis* takes place in *Nicomachean Ethics*, Book VI (1985). Although it has been suggested that Aristotle's remarks on *phronesis* are not always particularly illuminating, especially from a contemporary developmental and/or educational perspective, it does seem possible to derive a general account of *phronesis* from those texts that emphasies its diverse *functions* – hence refining and concretising 'the standard model'. The best way to convey the nature of those functions in contemporary psychological language is to say that the construct is made up of various (inter-related) *components*. We assume in what follows, in line with previous writings (Darnell *et al.*, 2019; Kristjánsson *et al.*, 2021b), that these are four.

The four-componential version of the 'standard model', which we have called the Aristotelian *Phronesis* Model (APM), constitutes a *pragmatic* hypothesis. It does not aim at unearthing essential structures of the human mind. The aim is simply to identify what roles *phronesis* is called upon to perform and how those can best be characterised for *explanatory* purposes and, subsequently, for purposes of *development and measurement*. Moreover, the components do not refer to psycho-moral capacities that are completely independent of one another and can be turned 'up' or 'down' in isolation; rather, they are inter-related as explained below (see further in Kristjánsson and Fowers, 2024, chap. 2).

Constitutive Function/Component
Phronesis involves the cognitive discriminatory ability to perceive the ethically salient aspects of a situation and to appreciate these as calling for specific kinds of responses. In the *phronimoi* (people possessing *phronesis*) this becomes a cognitive excellence in that, after having noted a salient moral feature of a concrete situation calling for a response, they will be able to weigh different considerations and see that, say, courage is required when the risk to one's life is not overwhelming but the object at stake is extremely valuable; or that honesty is required when one has wronged a friend. We could also refer to this function as *moral sensitivity or moral perception,* in order to link it more directly to the standard moral psychology/education literatures.

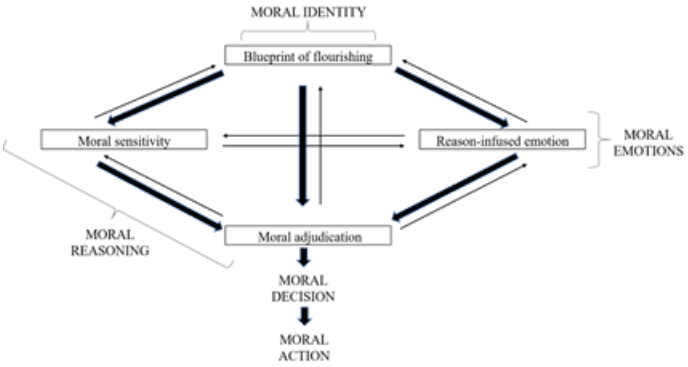


Figure 1: A Neo-Aristotelian Model of Wise (Phronetic) Moral Decision-Making.

Emotional Regulative Function/Component
Individuals foster their emotional wellbeing through *phronesis* by harmonising their emotional responses with their understandings of the ethically salient aspects of their situation, their judgement and their recognition of what is at stake. This is both because they will have already acquired habituated virtues, that is, have shaped their emotions in ways that motivate them to behave as a virtuous person would, and also because having formed these habits and consolidated them through understanding and reasoning, they will have a robust intellectual basis for them. For example, a *phronimos* might recognise that her appraisal of the situation is problematic, giving rise to an emotional response that is inappropriate to the situation. The emotion-regulative function can then help her adjust her emotion by, for instance, giving herself an inner 'talking to' or asking herself questions about what is prompting the ill-fitting emotional response. For this reason, we can also refer to this function, in a more standard Aristotelian way, as infusing emotion with reason.

Blueprint Function/Component.
The synthesising work of *phronesis* operates in conjunction with the agent's overall understanding of the kinds of things that matter for a flourishing life: the agent's own ethical aims and aspirations, her understanding of what it takes to live and act well and her need to live up to the standards that shape and are shaped by her understanding and experience of what matters. This amounts to what we call a blueprint of flourishing. A 'blueprint' has more similarity to what psychologists call 'moral identity' than a full-blown theoretical outline of the good life. *Phronetic* persons possess a general justifiable conception of the good life (*eudaimonia*) and adjust their overall reactions to that blueprint, thus furnishing it with motivational force. This does not mean that each ordinary person needs to have the same sophisticated comprehension of the 'grand end' of human life as a philosopher might have, in order to count as possessing *phronesis*. Rather than being an 'elite sport', the sort of grasp of a blueprint of the aims of human life informing *phronesis* is within the grasp of the ordinary well-brought-up individual. It draws upon the person's own life as a whole and determines the place that different goods occupy in the larger context.

Integrative Function/Component
Assume that we have identified a moral problem correctly as one potentially requiring input from two apparently conflicting moral virtues. Let us further assume that we have infused our relevant emotions with reason and that they are not obstructing the decision process. Finally, assume that we have a clear, non-self-deceptive identity of who we want to be – a blueprint of the good life – and an overall motivation to bring our reactions into line with that identity. That leaves just the final component of the four-componential construct: the integrative one – which we could also call its adjudicative function or, in line with standard moral psychology, denote as a form of 'moral reasoning'. Through this component, an individual integrates different virtue-relevant considerations, via a process of checks and balances, especially in circumstances where different ethically salient considerations, or different kinds of virtues or values, appear to be in conflict and agents need to negotiate dilemmatic space.

Figure 1 illustrates the overall conceptualisation of *phronesis*. Notice that we try to couch the components there in a language that will be more familiar to social scientists (entirely capitalised words) than the names of the four 'functions'.

2.4 THE UPSHOT OF THE JUBILEE CENTRE'S FIRST *PHRONESIS* REPORT

The neo-Aristotelian model of *phronesis* explained in Section 2.3 owes its inception to the first Jubilee Centre *Phronesis* Report (Kristjánsson *et al*., 2020) and a conceptual paper that was written in parallel and has since been widely cited (Darnell *et al*., 2019). A subsidiary aim was to contribute to the ongoing debate in moral psychology regarding the so-called gappiness problem: about what bridges the gap between moral knowledge and action. Not only did the first report offer theoretical guidance but it also described two empirical pilot studies (one conducted with an adult sample and the other with an adolescent sample) to test this model via a newly designed *phronesis* measure (LPM). The studies were conducted to investigate whether the proposed *phronesis* model is a suitable frame through which to investigate the relevant features of morality and their relation to prosocial behaviour.

In short, the findings of these pilot studies (which have since been written up in more detail for a peer-reviewed outlet: Darnell *et al*., 2022) were very positive:

- In both the pilot studies it was found, through structural equation modelling, that the hypothesised *phronesis* model fitted the data well. Previously validated measures that were predicted to be acceptable approximations of the components of the *phronesis* model were found to structurally relate to the predicted latent components in all but one case.
- Most importantly, the latent components were found to be structurally related to a predicted latent *phronesis* variable and, promisingly, this variable was found to predict the latent prosocial behaviour variable.
- Furthermore, the findings also suggested that the proposed *phronesis* model might have validity in both adult and adolescent samples, which has important implications for the 'gappiness problem' mentioned above.
- Although the model fitted both age groups, there was good evidence of developmental change, which is what one would want to see in a model that tracks a virtue-in-progress and is meant to have value in assessing educational interventions.

In terms of limitations, both studies relied on convenience samples and on available off-the-shelf measures that only approximated the functions of *phronesis*.

The authors, therefore, warned against too radical conclusions being drawn from those first results. While constituting a proof-of-concept validation, the need for more extensive replication studies was recorded.

2.5 THE UPSHOT OF THE JUBILEE CENTRE'S SECOND *PHRONESIS* REPORT, AND THE IDENTIFICATION OF PRACTICAL PROBLEMS

The second *Phonesis* Report (Kristjánsson *et al*., 2021a) did not add further data, as access to participants was severely limited during the Covid-19 pandemic. Instead, the research team re-analysed the existing data in light of various hypotheses derived from previous moral psychology literatures, such as:

- Female participants outperform male participants;
- Adult participants outperform adolescent participants;
- Female participants have higher levels of correspondence between chosen actions and justifications than male participants;
- Adult participants have higher levels of correspondence between chosen actions and justifications than adolescent participants;
- High correspondence between chosen actions and justifications predicts self-reported prosocial behaviour.

Some of those hypotheses were confirmed:

- Female participants outperformed male participants on all components of phronesis, and also on all measures targeting assumed sub-components;
- Adults outperformed adolescents on moral reasoning qua moral adjudication;
- Adults outperformed adolescents on moral emotion;
- Adults had higher correspondence between chosen actions and justifications than adolescents;
- Higher action–justification correspondence predicted two kinds of self-reported prosocial behaviour.

However, other findings clashed with expectations:

- Adolescents outperformed adults on moral perception and, indeed, on all its three presumed sub-components;
- There was no statistically significant difference between moral identity levels in adolescents and adults;
- While females had higher correspondence between chosen actions and justifications than males, the difference was not statistically significant.

The report offered some considerations to explain those findings, but otherwise repeated the call for a need to replicate the original pilot studies with different, larger samples – and also to use the *phronesis* instrument to measure the impact of actual *phronesis* interventions, pre- and post-.

> WE MUST THEREFORE SURVEY WHAT WE HAVE ALREADY SAID, BRINGING IT TO THE TEST OF THE FACTS OF LIFE, AND IF IT HARMONISES WITH THE FACTS WE MUST ACCEPT IT, BUT IF IT CLASHES WITH THEM WE MUST SUPPOSE IT TO BE MERE THEORY
>
> *Aristotle, Nicomachean Ethics, 1179a20-23*

Another project run within the Jubilee Centre at the time, about developing the practical wisdom of police-science students, offered a unique opportunity to try out the Jubilee Centre measure, pre- and post- (Kristjánsson *et al*., 2022). However, as it turned out very few of the participants had the patience to fill in the instrument before the intervention and even fewer afterwards, hence rendering findings statistically non-significant. As the lecturers acting as gatekeepers to the students explained, a 45-minute instrument – especially to be completed twice – was simply too tiring and unwieldy for the participants. An analysis of the kind of students who did, and did not, complete the instrument yielded some interesting results, however. Those students most motivated to pursue the *phronesis* intervention (as judged by their willingness to complete the time-consuming post-survey) scored significantly higher on measures of (i) the virtue aspect of a self-worth sub-measure and a sub-measure of moral self-relevance. Hence, sustained participation in the intervention was already predicted by some of the components that the intervention was meant to improve. The 'motivated students' also scored higher in perspective taking, empathic concern, prosociality, conscientiousness, extraversion and, most of all, agreeableness (Kristjánsson *et al*., 2022).

Over the course of the last couple of years, various external bodies, for instance universities teaching professional ethics, have contacted the Jubilee Centre and asked for permission to use the original LPM. When that has been done, users have complained both about the time it takes to complete the measure and how complicated it is to score: namely, it cannot be scored without drawing on expertise from members of the Jubilee Centre itself. It became ever more apparent, therefore, that in addition to the need for replication studies, a simpler version of the measure would need to be created: what psychologists often refer to as a 'quick and dirty' version. This explains the provenance of the additional studies conducted and described in the present report.

2.6 OVERALL EVALUATIVE GOALS

The initial motivation behind the *Phronesis* Project was to investigate how young people learn to bridge the gap between virtue literacy and moral reasoning about their virtues, on the one hand, and virtuous moral action, on the other. Aristotelian and neo-Aristotelian virtue ethics, which forms the theoretical basis of work in the Jubilee Centre for Character and Virtues (2022), has long assumed that the gradual development of the intellectual virtue of *phronesis* (or practical wisdom) in young people plays a fundamental role in the bridging of this moral 'gap', in particular as a means of adjudicating potential virtue conflicts. However, this assumption, although robust and respectable philosophically, had previously been underexplored psychologically and educationally.

The Jubilee Centre has come a long way, since the beginning of the *Phronesis* Project in 2019, in developing a nuanced conceptualisation and measure of *phronesis*. However, as explained above, the research journey has been paved with various obstacles, such as Covid-19, and practical implementation problems. The overall evaluative goals of this report, therefore, remain the same as in the previous two reports: namely, to assess the viability of the Jubilee Centre's model of *phronesis* (APM) both as a theoretical and psychometric construct. The research at issue puts Aristotelian moral psychology truly to the test: Is Aristotle's concept of *phronesis* an empirically viable notion, or is it a high-flown philosophical fiction, as critics of Aristotle-inspired moral psychology might suggest?

In addition to moving the practical discourse on *phronesis* forward within moral psychology and moral education, the present researchers hope that the findings reported on below also offer further general enlightenment on the role of practical wisdom in the good – flourishing, virtuous and well-rounded – life.

> FOR IN THAT HE [SOCRATES] THOUGHT ALL THE VIRTUES ARE [INSTANCES OF] *PHRONESIS*, HE WAS IN ERROR; BUT IN THAT HE THOUGHT THEY ALL REQUIRE *PHRONESIS*, HE WAS RIGHT
>
> *Aristotle, Nicomachean Ethics, 1144b18-21*

3 Methods

Various *phronesis*-related studies were conducted in the summer of 2023, four of which are reported upon in this report, as being most relevant to the development of the SPM.

3.1 STUDY 1: EXPLORATORY FACTOR ANALYSIS

3.1.1 Rationale

While previous studies conceptualised and tested the APM from the top down through comfirmatory factor analysis (CFA), Study 1 of this report was designed to devise a new measurement model of the APM from the bottom up using exploratory factor analysis (EFA) without *a priori* theoretical constraints beyond inspiring item generation.

3.1.2 Participants

We recruited a UK-representative sample of 1998 participants via a crowd-sourcing research participation website in the UK (*Prolific*). Their age, sex and ethnicity were checked against UK census data to ensure representativeness (see Table 1). All participants were paid £9 each for one hour of their time.

ONS Categories	% England and Wales	*N* (%) Current Sample
Sex [1]		
Female	51%	1024 (51.25%)
Male	49%	964 (48.25%)
Other	N/A	10 (0.50%)
Age [1]		
18–25	11%	253 (12.55%)
26–35	14%	367 (18.37%)
36–45	14%	388 (19.42%)
46–55	13%	333 (16.67%)
56–65	12%	444 (22.22%)
66–75	9%	186 (9.31%)
76–85	5%	25 (1.25%)
Ethnicity [2]		
White / White British	84.8%	1731 (86.64%)
Asian / Asian British	8%	142 (7.11%)
Black, African, Caribbean / Black British	3.5%	60 (3.00%)
Mixed / Multiple ethnic groups	1.9%	42 (2.10%)
Other		

Table 1. Demographics for participants in Study 1 versus demographics for the population for England and Wales per ONS Census data.

3.1.3 Measures

Given the length of the previous LPM, and how difficult users found it to score, we started afresh with a brand-new item list. These items were designed with the theorised four components of *phronesis* in mind. Items were initially generated using AI based on detailed descriptors of the construct of interest. These were iteratively refined in consultation with members of the research team to ensure we had adequate construct coverage. In some cases, it was possible to generate questions that could form parts of objective tests (e.g., being able to recognise whether a scenario bears moral relevance can have a 'correct' answer), and in other cases the construct of interest could only be subjectively self-assessed, meaning that self-reports were most appropriate (e.g., subjectively experienced moral emotions). An overview of these items can be found below.

3.1.3.1 Moral Perception

The question of whether one can correctly perceive morally salient aspects of different scenarios is not a subjective one. As such, we included two different kinds of objective tests (rather than self-report) of moral perception: recognising whether a situation is morally relevant at all, and identifying virtues at stake in those situations.

Situational Moral Relevance: We first developed 20 items in which participants were asked to discriminate whether a decision that a participant needed to make in a particular scenario would reflect on their character or not. Participants were presented with the following instruction: 'In the following section, you'll encounter a series of scenarios. Each one presents a situation that requires a decision. Your task is to determine which scenarios involve decisions that could impact your character. Reflect on these situations and consider the moral or ethical implications they might have on the individual involved.' This was followed by a scenario such as 'You've discovered a colleague is taking credit for your work but confronting them might cause tension in your team' (character-related according to a virtue ethical conception) or 'You're a book lover and have to decide which book to read next from a pile of equally appealing options' (not character-related according to a virtue ethical conception). Participants were given two response options: 'What I decide to do in this scenario *does not affect* my character' and 'What I decide to do in this scenario *affects* my character'. Participants' answers were then scored as being correct or incorrect.

Virtue Identification: The next set of moral perception items (15 in total) presented moral dilemmas in which participants were required to identify successfully which virtues were implicated in the scenario. Participants were instructed as follows: 'In this section, you will be presented with different scenarios, and four character traits that may or may not be relevant to the situation. Your job will be to select the two that you feel are most relevant to the situation.' They were then presented with a series of items in the following format 'You find a wallet on the ground with a significant amount of cash and the owner's identification. You need to decide what to do in this situation. Which of the following are most relevant to your decision? (select two answers from the following)'. Underneath were two relevant virtues from a virtue ethical perspective (e.g., honesty and practicality) and two less relevant virtues

(e.g., humour and resilience), from which participants were allowed to select two answers. As such, participants' scores varied from 0–2 for each question.

The items chosen under 'Moral Perception' were selected based on the same rationale as in the earlier *Phronesis* Project (Kristjánsson *et al*., 2020; Darnell *et al*., 2022) but are easier to score as no open questions were included this time.

3.1.3.2 Moral Identity

The question of whether a participant has a blueprint for their ideal moral self is, arguably, a subjectively assessable one. As such, for this component, we followed a traditional self-report survey format in which statements relevant to one's moral identity were presented (e.g., 'I believe my actions should reflect the type of person I aspire to be'), and participants responded on a five-point Likert scale (Strongly disagree – Strongly agree). As we had only one item format for this hypothesised *phronesis* component, we generated a larger array of items initially (25 in total).

The items chosen under 'Moral Identity' followed the format of standard moral identity questionnaires, such as those used in the earlier *Phronesis* Project (Kristjánsson *et al*., 2020; Darnell *et al*., 2022).

3.1.3.3 Moral Emotion

Moral emotion is also a subjective phenomenon; hence self-reports were deemed adequate for this component. Moral emotion is also multi-faceted, in theory. For instance, this could include the emotions people experience when they act morally or immorally or how well they can regulate their emotions. For this reason, we included multiple item formats under this *phronesis* component.

Emotional Reactivity: We first developed items to measure how somebody might feel if they acted morally or immorally in different scenarios. Participants were instructed as follows: 'Below are different scenarios in which you decide to take particular actions. If these were the actions you took in those scenarios, how would you feel about yourself?' This was followed by 20 statements, ten of which involved acting morally (e.g., 'A stranger drops a £100 note without noticing. You pick it up and return it to them'), and ten of which involved failing to act morally (e.g., 'You exaggerate a problem at work to damage a colleague's professional reputation'). Participants indicated how good or bad they would feel using a five-point Likert scale (Extremely bad – Extremely good).

Emotional Regulation: Although emotional regulation is partly subjective (e.g., personal affectedness) and partly objective (e.g., externalising behaviour, verifiable by others), we chose to use a self-report only as for assessment purposes it would neither be practical nor ethical to objectively measure emotional regulation in stressful scenarios. Participants were provided with the following instruction: 'Below, you will find several scenarios that might provoke an emotional response. For each situation, take a moment to reflect and imagine how you would typically react and then rate your ability to manage your emotions in that situation. By 'manage', we mean your ability to keep your feelings from overwhelming you and to maintain your composure. Please provide your answers on the following scale: 1 = Very poor ability to manage emotions; 2 = Poor ability to manage emotions; 3 = Neither poor nor good ability to manage emotions; 4 = Good ability to manage emotions; 5 = Very good ability to manage emotions. Please answer as honestly as you can.' This was followed by 20 scenarios, some of which involved being morally transgressed against (e.g., 'A stranger is rude to you in a public place for no apparent reason') and some of which involved everyday frustrations (e.g., 'You accidentally spill a drink on your clothes just as you are about to leave the house'). Participants were then asked to respond to each on a five-point Likert scale (Very poor – Very good).
We believe that the items selected under 'Emotional Regulation' better reflect the nature of the component in question than the off-the-shelf measures of empathy only that were used in the previous *Phronesis* Project (Kristjánsson et al., 2020; Darnell et al., 2022), as they target more directly the regulative element of *phronesis* rather than exclusively the general capacity to experience emotions.

3.1.3.4 Moral Adjudication

Moral adjudication involves the process of arriving integratively at a 'correct' moral choice as a result of proper deliberation. This involves two parts. The first is whether someone indeed selected the correct moral choice, as determined by some criterion. The second is to do with the process of making that choice in the right way.

Correct Moral Choices: Participants were provided with the following instruction: 'In the following section, you'll be presented with a series of scenarios. Each scenario has two main considerations that represent different potential responses to the situation. Your task is to decide how you would likely weigh up these considerations against one another if you were to respond in these scenarios. To do this, you must select one of seven boxes. The first and seventh box contain the two main considerations (see below). Selecting one of these two boxes means that you would focus completely on this consideration and not on the other at all. Selecting the boxes in between indicates that you would balance the two considerations to different degrees in deciding what to do. Please bear in mind that in real life, there might be more than just two considerations in these situations. However, for the purpose of this exercise, we ask that you do your best to make your decision based on the two considerations presented. There are no right or wrong answers in this task. We are interested in understanding your personal perspective and how you would handle these situations. Take your time and make sure to consider each scenario carefully before making your decision.' Following this instruction were 22 different items in the following format: 'A friend often vents their frustrations to you, but it has started to negatively impact your own mental health. Please use the scale below to express how you might weigh these considerations in determining your course of action.' This was followed by a 7-point scale. The first point on this scale involved a purely individualistic response (e.g., for the example item above, 'Cut off contact with the friend, preserving your mental health but leaving the friend without support'). At the opposite end of the scale was a purely prosocial response (e.g., 'Continue as is, supporting your friend but at the expense of your mental health'). Answers were scored in three different ways:

- In the first instance, answers were scored based on 4 being the 'correct' answer, representing an even balance between consideration of the self and others, in line with virtue ethical assumptions.
- In the second instance, the average flourishing scores (see below for measurement details) for each scale point were compared, and the scale point with the highest average flourishing score was defined as 'correct'. Then, participants were scored based on their answer's distance from that scale point for each question. For example, if the highest average flourishing was amongst people who scored '3' on the 7-point scale, then a person who selected '2' would be awarded a score of 1 (i.e., 1 scale point away from '3'), as would those who selected '4' (also one scale point away). This scoring

approach did not impose that the 'correct' answer is necessarily an even balance between individuality and prosociality; here it depends on the situation. A latent flourishing variable was chosen as the criterion for a 'correct' answer based on Aristotelian virtue ethical theory. As such, this scoring process deliberately created a dependency between this component and flourishing, in accordance with the underlying theory.

- Finally, answers were also scored on a 1-7 scale, assuming that higher scores representing greater prosociality would be 'good'.

In this way, we had 22 adjudication items scored three different ways each.

Moral Deliberation: It was not possible to measure moral deliberation using an objective approach so we asked participants to self-report on how they might solicit (e.g., 'Understanding all sides of a story helps me determine what's right') and check (e.g., 'I evaluate the reliability and credibility of the information sources before making a judgment') information pertinent to making a moral decision. 23 statements were presented to participants with which they were asked to agree or disagree on a five-point Likert scale (Strongly disagree – Strongly agree).

Moral Integration: Similarly, the most efficient way to tell whether someone integrates different *phronesis* components in their decision-making is to ask them. Therefore, participants were presented with 17 statements pertaining to moral integration with which they were asked to agree or disagree on a five-point Likert scale (Strongly disagree – Strongly agree). The following example item, 'In making a decision, I weigh my thoughts, feelings, and the situation at hand', shows us how moral emotions might be applied in a context-sensitive manner. Other items tap into different *phronesis* components (e.g., 'The person I want to be influences how I judge right from wrong' asks about alignment between Moral Adjudication and Moral Identity).

Adolescent Intermediate Concept Measure (AD-ICM; Thoma et al., 2013): The AD-ICM evaluates the moral thought processes of adolescents, focusing on their transition from self-centred to conventional thinking. Participants are presented with a story and a list of potential actions for the main character, which they must rate on a 5-point scale, from 'strongly believe this is a bad choice' to 'strongly believe this is a good choice' (e.g., 'Nikki should fire Beth, the weakest worker'). After this, participants select the top three most suitable and two least suitable actions. Next, they undergo a similar rating process for potential reasons the character might have for their actions (e.g., 'Beth knew she had to improve but did not'), rating each reason on the same 5-point scale. Participants then chose the top three most compelling and two least compelling reasons. Scoring for the AD-ICM is derived from participants' rankings and how their chosen items align with judgments made by an expert panel, as detailed in Thoma *et al.* (2013). High scores are given to top choices and reasons that experts deem acceptable and to the least favourable choices and reasons that experts consider unacceptable. Conversely, low scores are awarded to top choices and reasons labelled unacceptable by experts and to the least favourable ones that experts see as acceptable. Although the AD-ICM does not have enough cross-scenario variation to measure a situation-independent trait, this was included as it was part of the previous LPM. As the AD-ICM had proven too complicated to score for previous LPM users, Steve Thoma created a MS Excel file into which responses could be copied such that total scores for 'good moral choice' and 'good moral justification' could be computed using a series of built-in macros

3.1.3.5 Flourishing

The Well-Being Assessment (WBA; Weziak-Bialowolska *et al.*, 2021). The WBA is a tool devised to measure holistic well-being, drawing inspiration from a theoretical model that perceives human flourishing as a state where all life facets are positive. This method aligns with the World Health Organisation's definition of health, which encompasses not only mental and physical health but also the entirety of a person's being. Designed to evaluate well-being across six areas – emotional health, physical health, meaning and purpose, character strengths, social bonds, and financial stability – the WBA offers a relatively thorough insight into one's overall flourishing, and one that is in reasonable alignment with Aristotelian virtue theory. Although this is a self-report measure, its sub-scales have been found to predict more objective criteria (e.g., emotional health correlated with actual diagnoses of anxiety and depression, while physical health correlated with obesity, headaches, and migraines as assessed via insurance claim data).

We considered it vital, in line with Aristotelian virtue theory, to relate the adjudicative function of *phronesis* directly to flourishing, and also to tap into respondents' deliberation strategies. In general, the items chosen under 'Moral Adjudication' were more varied than in the earlier *Phronesis* Project (Kristjánsson *et al.*, 2020; Darnell *et al.*, 2022), which relied exclusively on the AD-ICM.

3.1.4 Procedures

All data were collected online, with participants completing the battery of questionnaires in a single one-hour session. The order of measures remained the same for each participant with questions completed in a random order within these measures in all cases apart from the AD-ICM, which must be completed sequentially. Informed consent was obtained for all participants prior to completing the questionnaire.

3.1.5 Analytic Strategy

As we used mostly new items in this study, we began with an exploratory factor analysis (EFA) to determine the number of factors represented by the items above on a purely empirical basis without predetermining the structure based on researcher preconceptions. Testing the previous LPM measurement model with a confirmatory factor analysis (CFA) in the first instance was not possible here as we used different items. EFA offered an alternative approach to delineating the factor structure of our largely new item set. In total, 189 items were included in the EFA. As such, we had greater than ten participants per item included in the model, as recommended in Carpenter (2018).

Following initial item reduction using EFA, the retained items and factors were entered into a structural equation model (SEM) and used to predict the criterion variable, a latent flourishing variable represented by the subscales of the WBA.

3.2 STUDY 2: CONFIRMING THE MODEL IN A UK SAMPLE

3.2.1 Rationale

Study 2 was designed to confirm the factor structure from Study 1 in a new UK-based sample. We also sought to establish criterion validity with a number of related measures.

3.2.2 Participants

This sample consisted of 1000 adults from the UK. This was also a representative sample based on ethnicity, age, and sex, reflecting the UK census data (see Table 2). Once again, participants were paid £9 for one hour of their time.

3.2.3 Measures

Data from 3.2.3.2–3.2.3.6 have not yet been analysed and will not be reported upon further in this report. However, we have outlined the other measures below for context.

3.2.3.1 The Short *Phronesis* Measure (SPM; Study 1)
We included only the items retained following the EFA in Study 1.

3.2.3.2 HEXACO-PI-R (Lee and Ashton, 2010)
This measure was used to measure The Big Five personality traits, Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, in addition to a sixth dimension that was considered especially pertinent, Honesty-Humility.

3.2.3.3 The Dark Tetrad (SD4; Palhaus *et al.*, 2020)
This measure was used to measure 'dark' personality traits that we might expect to be inversely related to morality. 28 items are used to measure Sadism, Machiavellianism, Narcissism, and Psychopathy, responded to on a five point Likert scale (Strongly disagree – Strongly Agree).

3.2.3.4 The Moral Foundations Questionnaire (MFQ-2; Atari et al., 2023)
The MFQ-2 was utilised to assess the presumed moral foundations rooted in Moral Foundations Theory. It distinguishes between the moral principles of Care, Loyalty, Authority, Purity, Equality and Proportionality.

3.2.3.5 Moral Relativism
We also included 20 items to determine levels of moral universalism (e.g., 'Even if everyone in a society believed an action was moral, it could still be objectively wrong') versus relativism (e.g., 'Different societies have different moral codes, and none can be said to be objectively correct').

3.2.3.6 Difficulty as Improvement (DAI; Yan et al., 2023)
Given that those high in *phronesis* might also seek to improve their characters though adversity, we also included the four-item DAI scale. These items (e.g., 'In a way, the difficulties I have today are strengthening my character to meet tomorrow's challenges') are scored on a five-point Likert scale (Strongly disagree – Strongly agree).

3.2.3.7 Propensity to Morally Disengage Scale (PMDS; Moore *et al.*, 2012)
The PMDS was used to assess the degree to which participants were willing to act immorally. This 16-item measure had eight factors related to distorting the consequences (e.g., 'Taking personal credit for ideas that were not your own is no big deal'), dehumanisation ('Some people have to be treated roughly because they lack feelings that can be hurt'), attribution of blame ('People who get mistreated have usually done something to bring it on themselves'), diffusion of responsibility ('It's okay to tell a lie if the group agrees that it's the best way to handle the situation'), displacement of responsibility ('People shouldn't be held accountable for doing questionable things when they were just doing what an authority figure told them to do'), moral justification ('It is alright to lie to keep your friends out of trouble'), advantageous comparisons ('Considering the ways people grossly misrepresent themselves, it's hardly a sin to inflate your own credentials a bit'), and euphemistic labelling ('It's okay to gloss over certain facts to make your point'). These items were answered on a seven-point Likert scale (Strongly disagree - Strongly agree).

3.2.4 Procedure

Participants were provided with an information page and were asked for their consent to take part. Participants then completed the questionnaires online, as before. Next, they were debriefed and thanked for their participation.

3.2.5 Analytical Strategy

Items retained from Study 1 were entered into a CFA to confirm the hypothesised factor structure in a new but comparable sample. As before, this CFA was well-powered with greater than ten participants per item included in the model. Next, a SEM was used to predict flourishing.

ONS Categories		% England and Wales	*N* (%) Current Sample
Sex [1]			
	Female	51%	508 (51.06%)
	Male	49%	482 (48.44%)
	Other	N/A	5 (0.50%)
Age [1]			
	18–25	11%	128 (12.86%)
	26–35	14%	176 (17.69%)
	36–45	14%	192 (19.30%)
	46–55	13%	162 (16.28%)
	56–65	12%	215 (21.61%)
	66–75	9%	110 (11.06%)
	76–85	5%	11 (1.11%)
Ethnicity [2]			
	White / White British	84.8%	874 (87.84%)
	Asian / Asian British	8%	63 (6.33%)
	Black, African, Caribbean / Black British	3.5%	30 (3.02%)
	Mixed / Multiple ethnic groups	1.9%	19 (1.91%)
	Other		

Table 2. Comparing demographic data from Study 2 to population demographics.

US Census Bureau Categories		% in the US	*N* (%) Current Sample
Sex [1]			
	Female	50.8%	501 (50.86%)
	Male	49.2%	469 (47.61%)
	Other	N/A	15 (1.52%)
Age [1]			
	18–25	9.5%	134 (13.60%)
	26–35	13.5%	175 (17.77%)
	36–45	12.7%	178 (18.07%)
	46–55	13.4%	170 (17.26%)
	56–65	12.6%	213 (21.62%)
	66–75	9.3%	101 (10.25%)
	76–85	4.5%	11 (1.12%)
Ethnicity [2]			
	White	57.8%	762 (77.36%)
	Asian	5.90%	54 (5.48%)
	Black or African American	13.4%	129 (13.10%)
	Two or more races	10.20%	31 (3.15%)
	Other (incl. Native American, Pacific Islands)	10.60%	19 (0.91%)

Table 3. Demographics for our US sample versus demographics reported by the US Census Bureau.

3.3 STUDY 3: TESTING THE MODEL IN A US SAMPLE

3.3.1 Rationale

Study 3 was designed to confirm the factor structure from Studies 1 and 2, but this time in a new US-based sample. We also sought to establish criterion validity with a number of measures related to moral disengagement.

3.3.2 Participants

This sample consisted of 1000 adults from the US. This was also a representative sample based on ethnicity, age and sex, reflecting the US census data (see Table 3). Once again, participants were paid £9 for one hour of their time.

3.3.3 Measures

As already noted, we have yet to report on associations between the SPM and the criterion variables outlined below. However, we have included them here for the sake of thoroughness and transparency. In this study, the criterion variables focused less on foundational moral attitudes, and more on how they might manifest in relation to practical concerns: namely, attitudes towards emerging technologies and consumer behaviour.

3.3.3.1 Select Measures from Study 2

As before, we used the HEXACO-PI-R to measure The Big Five personality traits plus Honesty-Humility. We also once again included the Dark Tetrad, as measured using the SD4.

3.3.3.2 Ethically Minded Consumer Behaviour Scale (EMCB; Riley and Kohlbacher, 2016)

This 26-item scale is designed to measure the extent to which consumers consider ethical issues when making purchasing decisions. The EMCB assesses various dimensions related to ethically-minded consumption, such as socially conscious consumption (e.g., 'I do not buy products which use advertising that depicts minority groups in a negative way') and environmentally conscious consumption (e.g., 'I have purchased products because they cause less pollution'). These statements are typically rated on a five-point Likert scale (Never true – Always true).

3.3.3.3 Attitudes Towards Emerging Technology

21 items were developed to understand participants' ethical positions on emerging technologies. Such is the pace with which technological innovation progresses that no validated scales existed for assessing moral positions in relation to technologies that are currently emerging. As such, this measure was designed by the research team to assess ethical positions in relation to Artificial Intelligence (e.g., Chat GPT), Autonomous Vehicles (i.e., self-driving cars), Blockchain technology (e.g., Bitcoin and Decentralised Finance), Bio-engineering (e.g., gene editing), Virtual Reality (e.g., Oculus or 'the metaverse'), Facial Recognition Technology (e.g., when logging into a laptop or phone) and Smart Devices (e.g., Alexa or smart watches). An example item is 'The benefits of Artificial Intelligence outweigh its potential risks.' to be answered on a five-point Likert scale from Strongly disagree to Strongly agree.

3.3.4 Procedure

Participants were provided with an information page and were asked for their consent to take part. Participants then completed the questionnaires online, as before. Next, they were debriefed and thanked for their participation.

3.3.5 Analytical Strategy

Items retained from Study 1 were entered into a CFA to confirm the hypothesised factor structure in a new but comparable sample. As before, this CFA was well-powered with greater than ten participants per item included in the model. Next, a SEM was used to predict flourishing.

3.4 STUDY FOUR: US-UK SAMPLE COMPARISONS

3.4.1 Rationale

Study 4 sought to test whether any of the *phronesis* components identified in Study 1, confirmed in a UK sample in Study 2, and confirmed in a US sample in Study 3, would differ between participants in the USA and UK.

3.4.2 Participants

We included the cleaned dataset from Study 2 (UK; n = 997), a cleaned dataset from Study 3 (US; n = 988), and a subset (US; n = 442) of the Study 3 data after removing outliers.

3.4.3 Analytical Strategy

We tested for differences in the ten sub-factors of *phronesis* identified in Study 1 between (i) the full UK sample and the abbreviated US sample and (ii) the full UK sample and the full US sample. For each of these analyses, the analysis code first loaded the three datasets into R, our statistical analysis programme, subsequent to which the ten variables of interest were identified within these datasets. The necessary variables were then extracted and copied into a new combined dataset. An extra column was added to identify which country each participant was from. From Study 3, we knew that the data were non-normal, so we used Mann Whitney U tests to identify differences, correcting for multiple tests using a Bonferroni correction. A rank biserial correlation was used to understand the magnitude of any differences found. Finally, overlapping density plots were used to visualise any group differences found.

3.5 ETHICAL CONSIDERATIONS

Ethical approval was granted for the research by the University of Birmingham Ethics Committee and informed consent was obtained for all participants.

> IT SEEMS PROPER, THEN, FOR A *PHRONETIC* PERSON TO BE ABLE TO DELIBERATE FINELY ABOUT WHAT IS GOOD AND BENEFICIAL ... NOT ABOUT SOME RESTRICTED AREA ... BUT ABOUT WHAT PROMOTES LIVING WELL IN GENERAL
>
> *Aristotle, Nicomachean Ethics, 1140a26-28*

WE MUST SEEK THE PROPER DEGREE OF EXACTNESS IN EVERY SCIENCE, SO THAT DIGRESSIONS DO NOT OVERWHELM OUR MAIN TASK

Aristotle,
Nicomachean Ethics, 1098a32-34

4 Findings

4.1 STUDY 1: EXPLORATORY FACTOR ANALYSIS

Data Preparation

Using R, our preferred statistical software, variables in the dataset were first changed to be numeric formatted variables, renamed, rescaled and scored. Ineligible participants were then excluded. Initially, we had a sample of 2073 participants including pilot participants. After filtering out participants who elected to withdraw, did not have a valid Prolific ID and those who had less than 95% survey completion, 74 participants were removed, leaving us with 1999 remaining.

Exploratory Factor Analysis

We conducted an EFA that included 15 Virtue Identification items, 4 different ICM total scores, 20 Situational Moral Relevance items, 20 Moral Emotion items, 20 Emotional Regulation items, 25 Moral Identity Items, 23 Moral Deliberation items, 17 Moral Integration items, and 22 Moral Adjudication items scored with flourishing as the criterion. KMO was .94, indicating adequate sampling. We also added a clause in the code not to run the EFA if Bartlett's test value was less than or equal to .05. We added another clause in the code to use a Principal Axis Factoring extraction method if skewness was > 2 or kurtosis was > 7; otherwise it would use Maximum Likelihood. The number of factors retainable was determined using parallel analysis. Finally, it was not clear to us from the outset whether factors should be correlated, especially given the potential for a multi-factor solution owing to a large sample size and item pool. Therefore, we used a Promax rotation method. Promax first applies an orthogonal Varimax rotation and then 'tilts' the axes to allow for obliqueness rather than seeking an oblique solution from the outset like an Oblimin rotation. This makes Promax computationally simpler and faster, especially when dealing with a large number of factors. For each of these factors, we suppressed loadings less than .4. In non-technical language, this means we conducted a detailed analysis to understand how different sets of questions or items group together, indicating underlying themes or factors. To ensure our analysis was accurate and relevant, we used various criteria and checks:

- First, we made sure we had enough data for the analysis (the KMO measure told us it was good to proceed).
- We did not run the analysis if Bartlett's test, which checks if our items are interrelated, failed to meet our set criteria.
- Depending on certain data characteristics (e.g., how skewed our data distribution was), we chose the best method to extract these themes or factors.
- We used a sophisticated method (parallel analysis) to decide how many such themes or factors we should consider, based on our total sample size.
- As for how these factors relate to each other, we started by assuming they were independent and then adjusted to account for possible relationships. This approach, using the Promax method, is like putting together a jigsaw puzzle by first laying out the pieces and then adjusting them to fit together. It is quicker and simpler, especially when dealing with many pieces or factors.
- Lastly, only the strong relationships or connections between items and factors (those with values above 0.4) were considered, ensuring we focused on the most meaningful associations.

Factor	Cronbach's Alpha	Retained	Example item	Factor name
1	.927	YES	I make sure to gather all the details before forming an opinion about what's right and wrong.	Moral Deliberation
2	.907	YES	I try to consider how my decisions today will reflect on the person I aspire to be in the future.	Identity Aspiration
3	.905	YES	(How able would you be to control your emotions if…) You accidentally spill a drink on your clothes just as you are about to leave the house.	Emotional Regulation
4	.829	YES	Consistency between my actions and my values enhances my self respect.	Moral Self-Relevance
5	.821	YES	My decisions are influenced by a mix of my thoughts, feelings, and personal beliefs.	Moral Integration
6	.833	YES	(How would you feel about yourself in these scenarios if these were the actions you took...) You exaggerate a problem at work to damage a colleague's professional reputation.	Negative Moral Affect
7	.821	YES	(Select the two most relevant virtues…) Your best friend reveals they cheated on an important exam and asks you to keep it a secret.	Virtue Identification
8	.762	YES	(How would you feel about yourself in these scenarios if these were the actions you took.?) A stranger drops a £100 note without noticing. You pick it up and return it to them.	Positive Moral Affect
9	.748	YES	(Does your choice in this situation reflect upon your character?) You're the manager of a restaurant and you catch one of your staff stealing food	Situational Moral Relevance
10	.725	YES	(Does your choice in this situation reflect upon your character?) You have an opportunity to invest in two promising startups, but you only have funds for one.	Situational Moral Irrelevance
11	.614	NO		
12	.393	NO		
13	.637	NO		
14	.431	NO		

Table 4: Internal reliabilities for factors retainable from parallel analysis with enough factor loadings greater than .4 to compute Cronbach's alpha.

In essence, we used a combination of advanced techniques to uncover and understand the main themes or factors from our large set of questions, ensuring our findings were both robust and meaningful.

In total, we extracted 17 factors using parallel analysis. The first 14 of these factors had enough items with loadings above .4 to compute Cronbach's alpha to assess their internal reliability. Ten factors had an acceptable internal reliability above .7, so these factors were retained (see Table 4).

4.2 STUDY 2: CONFIRMING THE MODEL IN A UK SAMPLE

Conceptually, the ten factors retained from Study 1 fitted with the theoretically informed four-component structure of *phronesis* as outlined in Table 5:

Moral Perception	Moral Emotion	Moral Identity	Moral Adjudication
Virtue Identification	Negative Moral Emotion	Identity Aspiration	Moral Deliberation
Situational Moral Relevance	Positive Moral Emotion	Moral Self-Relevance	Moral Integration
Situational Moral Irrelevance	Emotional Regulation		

Table 5: Mapping components from the UK EFA onto the theorised four components of Phronesis.

Therefore, we decided to conduct a hierarchical CFA in which our ten factors were considered to be sub-factors of the four conceptualised components of *Phronesis* (see Figure 1). This model generally fitted the data well (χ^2[45] = 2421.26, p<.001; *CFI* = .92, *TLI* = .876).

Next, we tested whether these four factors would load onto a superordinate *Phronesis* factor (see Figure 2). The model converged and showed a reasonable fit given its complexity (χ^2[32] = 304.759, p<.001; *CFI* = .89, *TLI* = .84).

In ordinary language, these findings mean that we explored if the four main themes or factors we identified can be grouped under one overarching theme, which we termed '*Phronesis*'. The results showed that our model, where the four factors combine under the umbrella of *Phronesis*, was a good representation of the actual data. While our statistics indicated that it

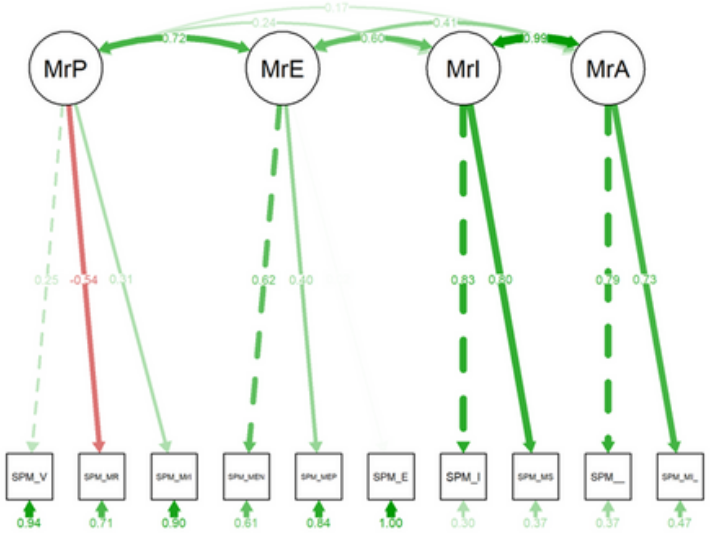


Figure 1: Testing whether our atheoretically derived components fit with the four theorised components of Phronesis.

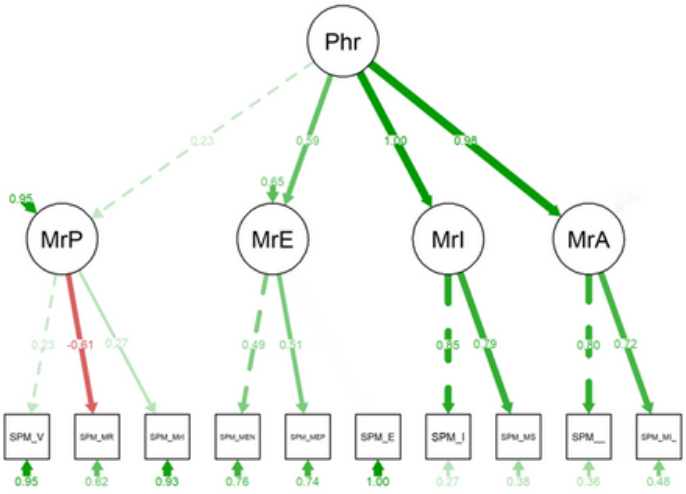


Figure 2: A four-component model of Phronesis in a UK sample..

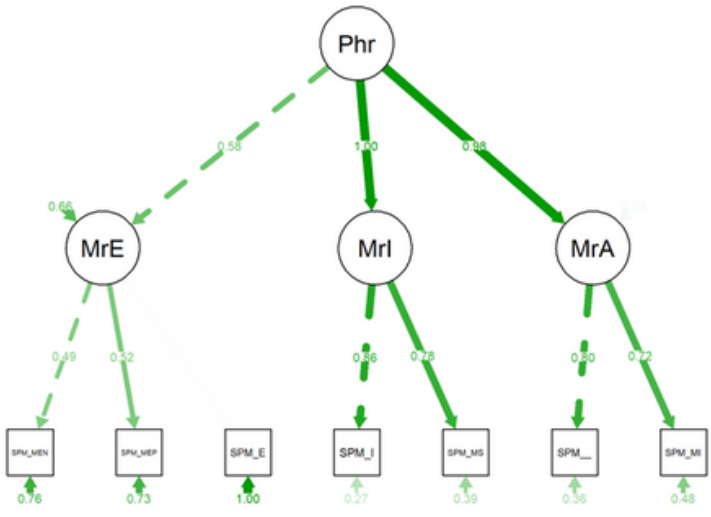


Figure 3: A three-component model of Phronesis in a UK sample.

was not a perfect fit, given the intricacy of the model, it was still a reasonably good fit, suggesting that *Phronesis* can be an overarching theme that encompasses the four factors we identified.

Next, given the notably lower association between Moral Perception and the overarching construct of *Phronesis*, on an exploratory basis, we excluded it in a further SEM, finding noticeably better model fit (χ^2 [21] = 153.510, p<.001; *CFI* = .94, *TLI* = .886; see Figure 3).

There are some theoretical reasons that could explain why moral perception would not fit the *Phronesis* construct as well as the other presumed components, and those will be explored in the Discussion section.

Finally, we compared the two and three-component models in terms of how well they predicted a latent flourishing variable. The three-component model (χ^2 [78] = 580.608, p<.001; *CFI* = .90, *TLI* = .87; see Figure 4) and the four component models (χ^2 [100] = 773.535, p<.001; *CFI* = .87, *TLI* = .85; see Figure 5), were similar. A direct comparison using the Chi-Squared Difference Test revealed a significant difference in the fits of the two models, with a chi-square difference of 192.93 and a highly significant p-value (p < .001). Moreover, both a lower Akaike's Information Criteria (AIC) and a lower Bayesian information criteria (BIC) value suggested that the three-component model had a better fit to the data than the four-component model. The Root Mean Squared Error of Approximation (*RMSEA*) value for the four-component model (.08) was lower than the three-component model (.09), showing that not all fit indices favoured the three-component model. Overall, the two models might be considered roughly equivalent, with one being more parsimonious and the other more faithful to the underlying theory.

Next, we sought to explore the criterion validity of our measure with the most conceptually important outcome measure: Flourishing. We computed our total scores based on the model identified in Study 1. Additionally, we created composite scores for each of the four APM components by averaging z-scores for the relevant sub-components. We then correlated our ten factors with different aspects of flourishing (see Table 6).

Interestingly, we found that Moral Perception and its components were not associated with flourishing. This may be because understanding what is at stake morally is not the same as wanting to do

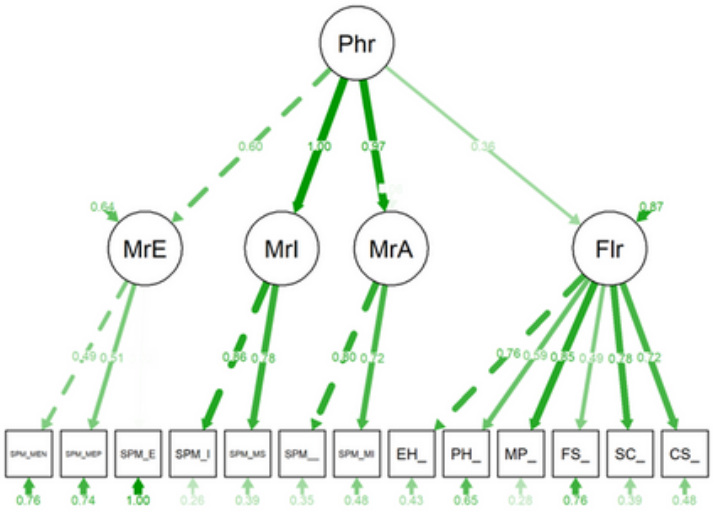


Figure 4: Predicting flourishing using a three-component Phronesis model in a UK sample.

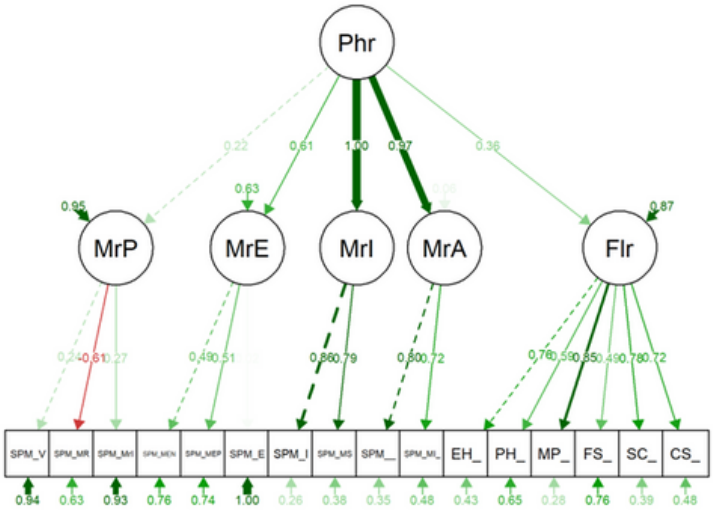


Figure 5: Predicting flourishing using a four-component Phronesis model in a UK sample.

	Emotional Health	Physical Health	Meaning and Purpose	Financial Security	Social Connectedness	Character Strengths
Moral Perception	.01	-.015	.059	.016	-.011	-.008
Virtue Identification	.01	-.004	.044	.031	-.001	.015
Situational Moral Relevance	.026	-.032	.012	.033	-.005	.004
Situational Moral Irrelevance	-.026	.018	.03	-.033	-.01	-.029
Moral Emotion	.237***	.207***	.302***	.027	.265***	.426***
Negative Moral Emotion	-.132***	-.14***	-.267***	-.073*	-.209***	-.316***
Positive Moral Emotion	.122***	.126***	.179***	-.026	.17***	.295***
Emotional Regulation	.185***	.137***	.139***	.01	.135***	.216***
Moral Identity	.192***	.173***	.336***	.077*	.243***	.429***
Identity Aspiration	.223***	.173***	.326***	.103**	.269***	.425***
Moral Self-Relevance	.115***	.117***	.256***	.032	.157***	.321***
Moral Adjudication	.188***	.191***	.285***	.072*	.207***	.355***
Moral Deliberation	.208***	.213***	.308***	.104**	.208***	.434***
Moral Integration	.109***	.098**	.156***	.015	.14***	.174***

Table 6: Correlations between Phronesis components and flourishing in a UK sample.

good, or indeed, doing good. Moral Emotion overall was positively related to all aspects of flourishing except Financial Security. This was true for the Emotional Regulation and Positive Moral Emotion aspects, while Negative Moral Emotion, which we reverse scored before computing the composite variable, was inversely related to flourishing. Moral Identity and its facets were especially strong predictors of Character Strengths and Meaning and Purpose, but overall tended to predict all aspects of flourishing. The coefficients were notably weak for Financial Security, however. Moral Adjudication and its facets were also associated with flourishing.

4.3 STUDY 3: TESTING THE MODEL IN A US SAMPLE

Preliminary Data Cleaning

As with the UK sample in Study 2, in Study 3's US-representative sample, we first loaded the data and removed participants with < 95% completion, those who did not have a valid Prolific ID, and those who failed one or more attention checks. Variables were renamed, rescaled and converted to a numeric format if applicable. Two items from the Moral Self-Relevance subscale were reverse scored.

Next, we computed Cronbach's alpha for the ten subscales. They were all found to be internally reliable in the US sample (.72-.91). Total scores were then computed for the ten subscales. We then averaged z-scores for these subscales in accordance with each of the four *Phronesis* components and created superordinate total scores for Moral Perception, Moral Identity, Moral Emotion, and Moral Adjudication.

SEM Modelling

Having found a hierarchical *Phronesis* factor in the previous study, we sought to test this in a US sample. Initially, the models would not converge. We therefore explored our data distributions to understand why. Since we had a large sample, normality tests were likely to be overly sensitive to small deviations from normality, so instead we plotted the ten composite scores. We found that, in the US sample (see Figure 9), all but the Emotional Regulation sub-factor seemed to be skewed. We therefore removed any cases where participants scored greater than 2 SD from the mean for any of the ten variables. This cut our sample size substantially from 998 to 442.

The three-component *Phronesis* model converged and showed good model fit that was almost identical to what we found in the UK sample (χ^2 [12] = 64.264, p<.001; *CFI* = .90, *TLI* = .83; see Figure 6).

When we used this model to predict flourishing, the model fit was still relatively high (χ^2 [62] = 276.756, p<.001; *CFI* = .88, *TLI* = .85; see Figure 7) and the superordinate *Phronesis* factor was shown to predict flourishing with a similar magnitude to that of the UK sample.

Next, we repeated the process for the four-component version. The overall four-component *Phronesis* model did not converge in the US sample. The four-component model that predicted flourishing (see Figure 8) did, however, showing reasonable model fit (χ^2 [120] = 1900.920, p<.001; *CFI* = .87, *TLI* = .84) and predicted flourishing with the same magnitude as in the previous study.

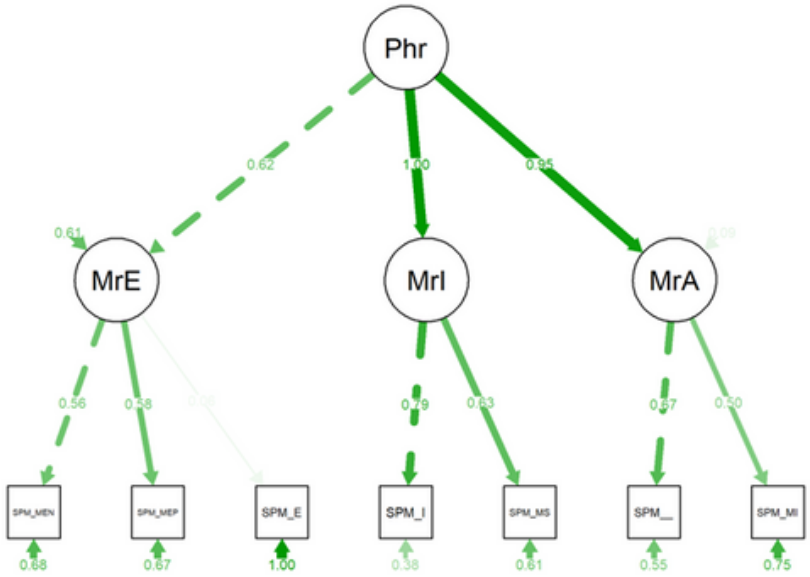


Figure 6: A three-component model of Phronesis fits the data well in a US sub-sample.

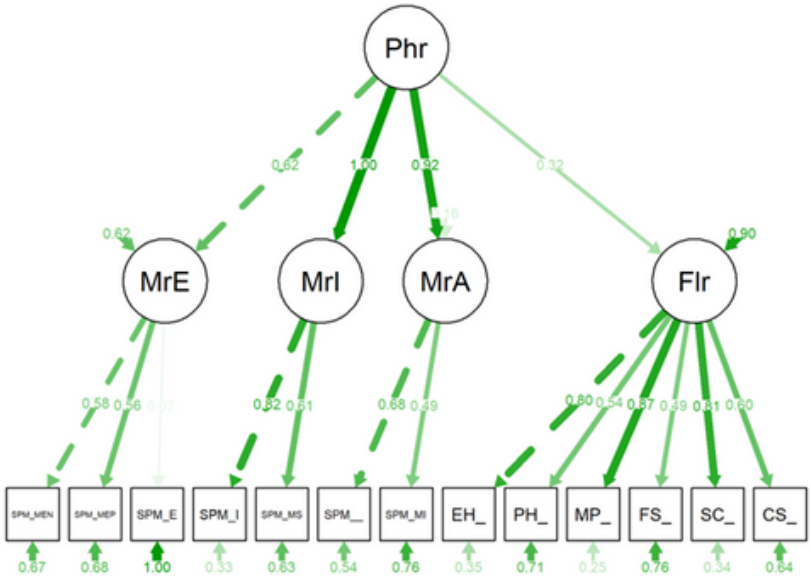


Figure 7: A three-component model of Phronesis predicts flourishing in a US sample.

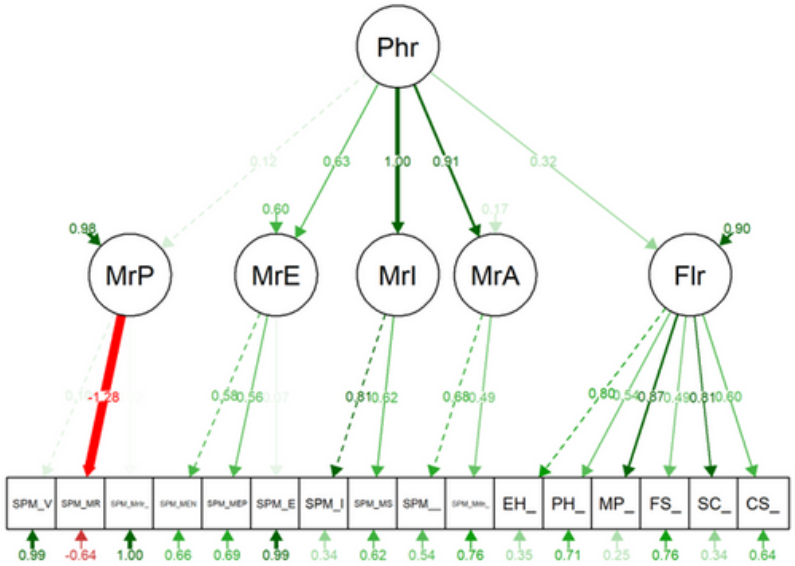


Figure 8: A four component model of Phronesis predicts flourishing in a US sample.

In comparing our two structural equation models, the three-component model was contrasted against a four-component model. The chi-squared difference test was statistically significant ($\Delta\chi^2(38) = 62.028$, $p = 0.008$), suggesting that the inclusion of the fourth component significantly improved the model fit. Furthermore, the RMSEA values provide additional insights: while the three-component model presented an RMSEA of .089, slightly exceeding commonly accepted thresholds for good fit (.08), the four-component model yielded an RMSEA of .074, indicative of a more acceptable fit. Though the AIC and BIC values for the three-component model were lower, suggesting a better fit when considering model parsimony, the improved RMSEA in the four-component model is also arguably compelling, given the underlying theory and acceptable internal reliability of the individual scales. Given these considerations, while the three-component model may seem preferable based on information criteria, the significant chi-squared difference combined with a more satisfactory RMSEA for the four-component model underscores the value of balancing statistical evidence with theoretical justification when deciding on the most appropriate model.

4.4 STUDY 4: US-UK SAMPLE COMPARISONS

Overall, it appears that the *Phronesis* model developed and verified in a UK sample can also be found in a subset of the US sample. However, the next study sought to establish where differences might lie in our key variables across the two countries. To do this, we conducted two comparative analyses. The first compared the Study 2 data (UK) with the final dataset from Study 3 (US) that excluded outliers of less than or equal to 2 SD. The second analysis did the same, but with the original non-trimmed Study 3 dataset.

UK vs US Trimmed

In the comparative analysis between UK and US participants across ten key measures, the results revealed significant differences for all variables, as indicated by their Bonferroni-adjusted p-values. Notably:

- Variables like Virtue Identification (rank biserial $r = .243$), Situational Moral Irrelevance ($r = 0.374$), Negative Moral Emotion ($r = .305$), Positive Moral Emotion ($r = .327$), Emotional Regulation ($r = .271$), Virtuous Identity Aspiration ($r = .476$), Moral Self-Relevance ($r = .545$), Moral Deliberation ($r = .436$) and Moral Integration ($r = .236$) displayed higher scores in the US compared to the UK.
- Conversely, the Situational Moral Relevance variable ($r = -.294$) indicated that UK participants scored higher than their US counterparts.

The effect sizes, represented by the rank biserial correlations, provide a measure of the magnitude of differences between the two countries on each variable. While all effect sizes are of practical significance given the large samples, it is worth noting that the Virtuous Identity Aspiration ($r = .476$) and Moral Self-Relevance ($r = .545$) variables exhibit notably large effect sizes, pointing to particularly strong differences between the two groups on these variables. In simple terms, it appeared that this subset of the US sample in which the SPM model fitted differed from the UK in that the US sample scored higher overall on all but one of the ten sub-components of *Phronesis* identified in Study 1.

UK vs US Full

In the comparative analysis between UK and US participants across ten key measures, the results revealed significant differences for all variables, as indicated by their Bonferroni-adjusted p-values. Notably:

- Variables like Virtue Identification (rank biserial $r = .139$), Situational Moral Irrelevance ($r = .127$), Negative Moral Emotion ($r = -0.091$), Positive Moral Emotion ($r = .163$), Emotional Regulation ($r = .199$), Virtuous Identity Aspiration ($r = .286$), Moral Self-Relevance ($r = .281$), Moral Deliberation ($r = .244$) and Moral Integration ($r = .125$) displayed higher scores in the US compared to the UK.
- Conversely, the Situational Moral Relevance variable ($r = -.120$) indicated that UK participants scored higher than their US counterparts.

While all effect sizes are of practical significance given the large samples, once again it is worth noting that the Virtuous Identity Aspiration ($r = 0.286$) and Moral Self-Relevance ($r = 0.281$) variables exhibit notably large effect sizes, pointing to particularly strong differences between the two groups on these variables. Again, in simple terms, the representative US sample showed the same differences across the ten sub-components of *Phronesis* as the non-representative sub-sample, but the differences were smaller.

It should finally be noted that we observed a differential model fit in the US sample compared with the UK sample. Overlapping density plots (see Figure 9) underscored this discrepancy, with the US participants generally showcasing a higher density of scores towards the edges of the respective distributions. Notably, as already spelled out, US participants recorded higher scores on the majority of the *Phronesis* subscales, suggesting some systematic difference between the two samples. This was not merely a difference in moral self-appraisal in US participants vs UK participants; US participants also scored noticeably higher on objective tests of virtue identification (see Figure 9, upper-left).

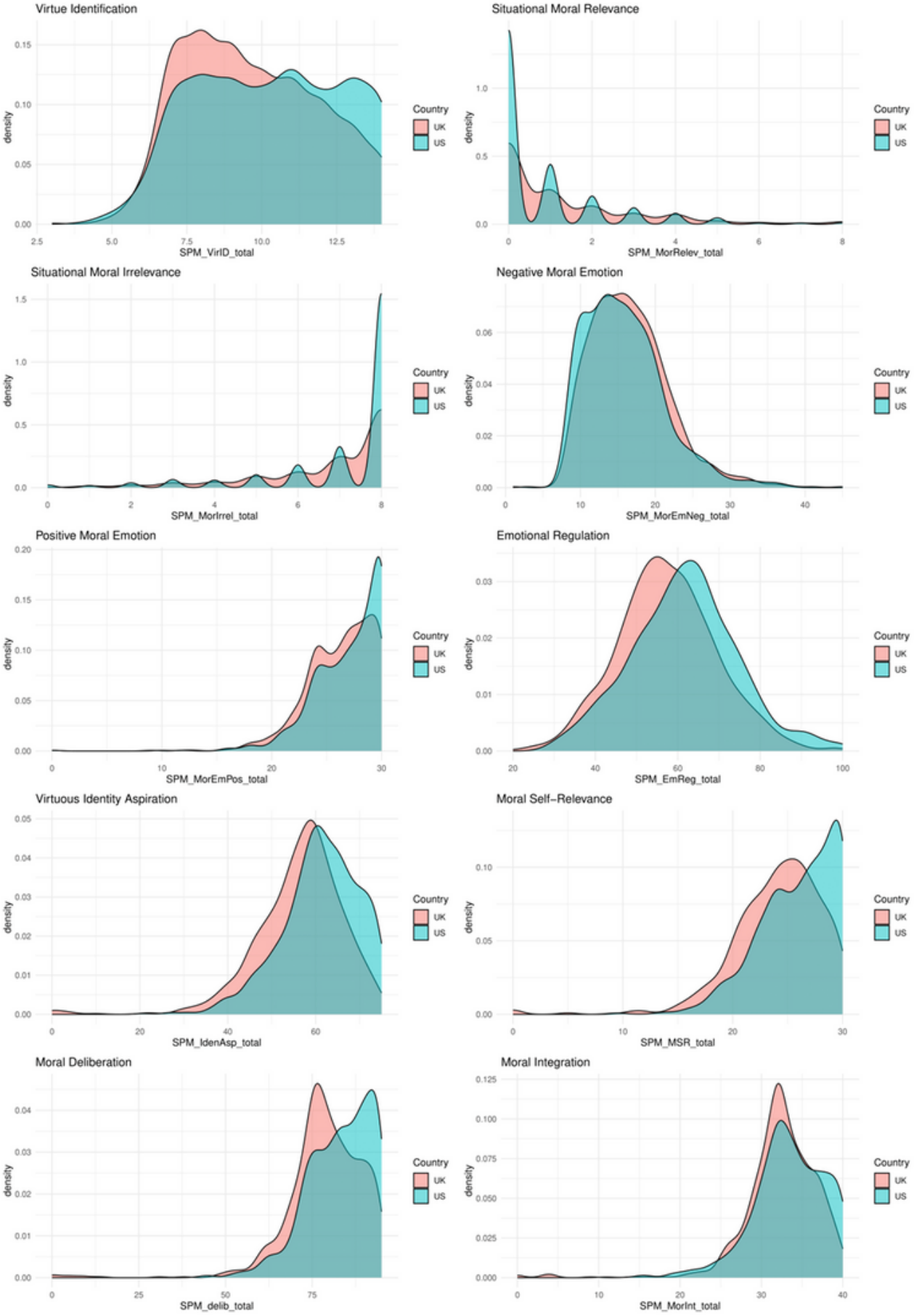


Figure 9: Density plots revealing that US participants more frequently score near the edges of the distribution.

5 Discussion

5.1 BUILDING ON THE LPM

An initial impetus behind the development of the Short *Phronesis* Measure (SPM) was to create a tool that was more pragmatic for broad research and application purposes than the LPM, while not compromising on its virtue theoretical assumptions. It is helpful to rehearse the research questions from Section 1:

- Does a new exercise in the design of a shorter *phronesis* instrument, using a greater number of participants of varied ages inside and outside of the U.K., and starting with an exploratory factor analysis, confirm the putative viability of the Jubilee Centre's *phronesis* construct?
- Does a hierarchical confirmatory factor analysis, in which the factors extracted from the exploratory factor analysis are considered sub-factors of the conceptualised components of *phronesis*, fit the data?
- Does the model extracted from the confirmatory factor analysis predict a latent flourishing variable?
- More generally, can the Jubilee Centre produce an instrument to measure *phronesis* that is shorter and easier to score than the earlier incarnation?

The first and second research questions contain within them various sub-questions. As the studies used to develop the LPM relied solely on UK respondents, it is worth commenting here first on the significant differences found between the US and UK samples in our current studies, with participants from the first scoring higher on all but one of the *Phronesis* variables. While surprising, this finding is not entirely so, as previous Jubilee Centre studies have identified significant cultural differences between UK and US samples (see esp. Morgan *et al*., 2014). Moreover, as many of our survey questions rely on self-reports, some people could see this finding as a confirmation of the standard archetypes of self-confident Americans versus self-effacing Brits. However, this would not explain the higher scores from US participants on more objective tests of moral perception (e.g., correct Virtue Identification). Moreover, all subscales of the SPM were found to have strong internal reliability in all samples.

While systematic cultural variance between the USA and UK (of whatever provenance) might be a contributing factor, it would be prudent not to leap to definitive conclusions based solely on this observation. The higher scores in the US sample might have a variety of explanations such as a more specific cultural difference in relation to degrees of moral polarisation, or even variations in how certain moral virtues or actions are perceived, understood and valued. It is also possible that contextual factors, such as prevailing societal narratives, recent events or broader educational emphasis on moral reasoning within the two regions, could have influenced the differential scoring patterns observed. Such differences further emphasise the intricacies of studying complex constructs like *phronesis* across diverse samples. They highlight the importance of understanding the specific biopsychosocial environments within which a measure is applied, as this can affect the interpretation of the measure.

While understanding these discrepancies is essential, the true merit of a measure lies not only in its theoretical soundness but also in its applied utility. As a tool designed to gauge and predict certain outcomes or behaviours, the utility of the SPM becomes paramount. Thus, future research should prioritise evaluating how well the sub-measures of *phronesis* predict relevant outcomes across various contexts (cultural and otherwise), ensuring the measure functions as a versatile and practical instrument in the study of moral virtues. Even if it turns out that *phronesis* scores do differ

systematically between cultures or sub-cultures, such a finding would not undermine the usefulness of the SPM, since we envisage that its most common use will be to pre-and post-test the same cohorts before and after a given intervention. In such cases, what matters is progress from T1 to T2 in the given cohort, rather than differences between cohorts.

However we choose to interpret the cultural variance found in our studies, the large size and representativeness of our samples must be considered among its strengths. While the LPM, in its nascent stages, employed convenience samples primarily for proof-of-concept, our approach aimed to build on this and capture a more genuine cross-section of the UK and US populations. By securing (relatively) representative samples from these regions, we not only fortified the reliability and validity of the SPM but also positioned it as a measure that can truly generalise its findings to broader populations. This representativeness elevates the external validity of our results, ensuring that the inferences drawn from our research can more reliably be extended and applied to the wider populace of the UK and US, thus offering more robust insights into the nature of *phronesis* in these societies. A note of caution, however: While the UK samples were broadly representative of the general UK population (see Tables 1 and 2), the US sample was less representative of the US population than one might expect based on the US Census Bureau statistics, with White participants oversampled via Prolific (see Table 3).

The LPM marked a significant advancement in the empirical study of *phronesis*, successfully affirming a four-factor structure consistent with Aristotelian theory. This validation underscores the LPM's foundational role in supporting theory with empirical evidence in this domain. However, the chosen methodology (Darnell *et al.*, 2022) directly applied structural equation modelling (SEM) without a preceding phase of exploratory factor analysis (EFA). While the results yielded a good model fit in a later sample, employing EFA after theoretically informed item generation, as in our present studies, offers an advantage. This step ensures the data's natural structure is identified without imposing theoretical biases prematurely. This process enabled an organic and bottom-up unveiling of the data structure that could later be reconciled with the top-down APM theory.

One noticeable limitation of the LPM was its confinement to just two distinct moral scenarios. Although these scenarios were investigated in significant depth, providing valuable insights, they potentially constrained the measure's capacity to generalise about the person across diverse situations. Such a focus makes it inherently challenging to abstract a consistent trait-like quality of *phronesis*. Cross-scenario variation is a crucial element for trait measures, which aim to capture stable characteristics about an individual regardless of specific situational nuances. In response to this challenge, our SPM introduced a broader spectrum of moral dilemmas, capturing a more diverse and comprehensive range of thoughts and behaviours. This expansion not only allows for a robust trait inference but also ensures that the measure remains sensitive to the varied ways in which *phronesis* can manifest in different contexts.

The eventual identification of a three-componential *Phronesis* construct as the best-fit model may indicate that the current studies were not as supportive of the Jubilee Centre's four-componential model as the studies underlying the LPM. However, there are various considerations that mitigate that conclusion. First, the four-componential model of Aristotelian *phronesis*, or APM as it has been called (Kristjánsson and Fowers, 2024), was put forward as a pragmatic model, as explained in Section 2.3. It was never meant to 'carve nature at its joints'. Krettenauer (2019) may be right that, from a structural point of view, as well as the point of conceptual parsimony, two of the components identified in the model might better be seen as preconditions rather than constituents of *phronesis*, which would leave two essential components only: the constitutive and integrative ones. Nevertheless, from a pragmatic perspective – as the two 'preconditions' are also necessary for *phronesis* to function – they were included in the APM as components. Aristotle himself was, indeed, notoriously ambiguous about some variables in his core concepts, such as *eudaimonia*, for example sometimes speaking of good friends and good health as preconditions of *eudaimonia* but sometimes as constituents. More generally, nothing precludes the same item x from being seen simultaneously as instrumentally (precondition) and intrinsically (component) related to y, when looked at from different perspectives.

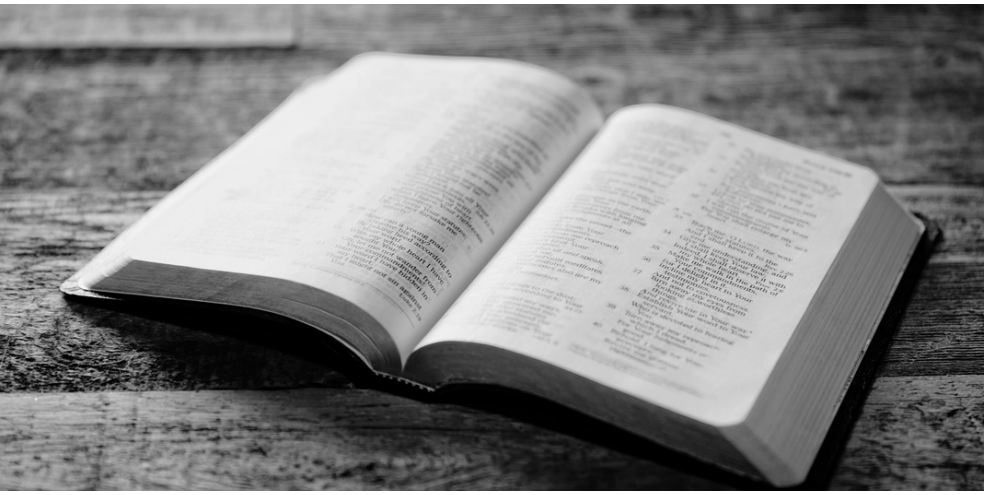
In our current studies, three of the four proposed factors did, in fact, have an excellent fit with a general *phronesis* construct. The odd one, out, moral perception (aka the constitutive component), is not one of the factors that Krettenauer (2019), for one, questions as integral to *phronesis*. It is also possible that perception has a worse fit with *phronesis* in the current studies than in the studies undergirding the LPM, because the participants in previous studies were on average much younger, and it could well be that, developmentally, general virtue literacy (understanding virtue terms and noticing virtues in situations) is a more integral developmental task for younger than older people. A closer look at age differences in our samples may possibly help answer that question. More generally speaking, perception may function more as a precondition of moral deliberation than an integral part of it. One could argue that evil people also need to spot the relevant virtues and vices involved in a situation in order to figure out the best way to pursue their evil ends. In that sense, the relationship between having a keen sense of the moral details of the situation does not possess the same *phronetic* credentials as, for example, a proper moral identity (*qua* the blueprint component) does.

Second, recall that the finding about the superiority of a three-component model was not unambiguous psychometrically. Though the AIC and BIC values for the three-component model were lower, suggesting a better fit when considering model parsimony, the improved RMSEA in the four-component model is also arguably a consideration, when complemented with the theoretical arguments that support a four-component model. Therefore, the emergent structure is reconcilable with the theorised four-component model. This reminds us of the fact that statistical modelling, based on model parsimony, only takes us so far. If there are compelling reasons for preferring an alternative model from a theoretical perspective, statistics – especially when those do not yield a fully unambiguous conclusion – do not necessarily carry greater weight. All in all, the theorists behind the APM will take great comfort in the fact that both the studies underlying the LPM and the SPM support the APM in all essentials.

The third research question is relevant for present purposes, because it directs our attention to the most conspicuous novelty of the current studies, compared to the earlier ones: namely, the focus on the relationship between *phronesis* and flourishing – an association that would, for Aristotle, have mattered most, as flourishing (*eudaimonia*) is nothing less than the grounding concept of his whole virtue ethics. Importantly, the three

phronesis factors that best fitted the model predicted all aspects of flourishing strongly (apart from the aspect of financial security). In contrast, the LPM's primary focus was on prosociality as an outcome. Though valuable, this perspective might be too narrow when considering the broader theoretical context, which emphasises flourishing as the paramount criterion. The SPM directly addresses this by aligning with the comprehensive and validated Well-Being Assessment (WBA), shifting the focus to flourishing as a much wider concept that incorporates both subjective (e.g., meaning and purpose) and objective flourishing (e.g., physical health). This comprehensive view of flourishing is, as already indicated, more closely aligned with the overarching Aristotelian theory than the concept of prosociality, adding considerable backbone to the viability of the SPM and the APM underpinning it.

Finally, the fourth research question addresses the most significant practical issue: Have we created an instrument that is sufficiently short, reliable, valid and easy to score? Arguably, the answer to all those questions is in the affirmative. Recall that the LPM (Darnell *et al.*, 2022), though pioneering, presented certain practical challenges given its length and intricate scoring mechanism. In contrast, the SPM is comparatively brief and has a simplified scoring process. It should not take more than 20 minutes to complete, on average, with many of our participants completing the measure even faster than this. Our studies intentionally designed the SPM to be concise, ensuring that participants could complete it in a shorter span of time without compromising the integrity and depth of the information collected. The more streamlined scoring mechanism also ensures that researchers and practitioners can quickly interpret the results, making the SPM a highly scalable and versatile tool, apt for diverse research contexts and large-scale studies.



Although the new SPM will not be made publicly available until all the large data set, accumulated in the summer of 2023, has been thoroughly analysed and subjected to academic peer-review, exploratory uses of the new measure will be allowed after the publication of this report (see information at the end of the Executive Summary).

5.2 SITUATIONAL AFFORDANCES AND FURTHER RESEARCH

The SPM, having been developed with representative samples from both the USA and the UK, informs us about the structure and measurement of *phronesis* in these countries in general, advancing our foundational theoretical understanding. However, it remains to be seen to what extent these identified components will be important within more specific situations. This echoes the Jubilee Centre's previous focus on accounting for situational affordances. For instance, Harrison et al.'s (2023) attempted to measure four hypothesised components of cyber-wisdom that were initially derived from the APM/LPM and then took account of domain-specific knowledge (e.g., cyber-wisdom literacy). Future research might explore the predictive validity of the SPM in new contexts such as virtual environments (e.g., online or in VR) or at work, or whether those might be better gauged with more domain-specific measures. Such measures may also be important to use alongside the SPM moving forward. For instance, the ICM has been used to test whether a person's responses align with what an 'expert' in a particular domain (e.g., the ICM for teaching; see Kerr, 2021) deems to be correct in a moral dilemma scenario. Scores on such measures can then be used to predict subsequent behaviour or attitudes of interest (e.g., for teaching, it could be burnout, student satisfaction, etc.). However, by controlling for trait-level practical wisdom using a tool like the SPM, it would be possible to test whether domain-specific measures (e.g., the teaching ICM) predict outcomes within the domain of interest (e.g., student satisfaction) over and above the SPM. In other words, we could separate out how much of a behaviour in context is accounted for by trait *phronesis* (i.e., something about the person per se) versus practical wisdom specifically in relation to that context (cf. also Grossmann, 2017).

Prior to such work on domain-specific measures, the Jubilee Centre plans to continue analysing the data set gathered for the present studies. Age differences may, for example, reveal the developmental value of the measure. Also worth mentioning are prospective findings from the subsidiary measures that were listed in Section 3 but have not been analysed yet. Interesting research questions that await answers are, for example:

- How is the SPM related to trait measures of personality, including the most morally relevant personality trait, Honesty-Humility?
- How is the SPM related to the four 'dark' personality traits: Machiavellianism, Psychopathy, Narcissism and Sadism?
- How is the SPM related to constructs within adjacent moral theories (e.g., Moral Foundations Theory)?
- How is the SPM related to variables related to the abdication of moral responsibility?
- How is the SPM related to moral positions on day-to-day matters (e.g., ethical consumer behaviour or ethical adoption of emerging technologies)?
- Are there meaningful demographic differences in the SPM scores?

5.3 CONCLUDING REMARK

In summation, while we believe that the introduction of the SPM is a worthwhile advancement in the study of virtue ethics in general and practical wisdom in particular, it marks the beginning of a new chapter rather than a conclusion. There is both a need and an opportunity for researchers to approach the SPM and its applications with a balance of optimism and critical evaluation. It is through such balanced inquiry that the model's robustness, relevance and potential to genuinely advance our understanding of practical wisdom will be realised. The coming years, marked by further research and exploration surrounding the SPM, promise a wealth of knowledge and insights into the intricate realm of *phronesis* and human flourishing.

References

Aristotle (1985) *Nicomachean Ethics*, trans. T. Irwin, Indianapolis: Hackett Publishing.

Ashton, M.C. and Lee, K. (2010) 'Trait and Source Factors in HEXACO-PI-R Self-and Observer Reports', *European Journal of Personality*, vol. 24, no. 3, pp.278-289.

Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T. and Dehghani, M. (2023) 'Morality beyond the WEIRD: How the Nomological Network of Morality Varies across Cultures', *Journal of Personality and Social Psychology*, https://doi.org/10.1037/pspp0000470

Carpenter, S. (2018) 'Ten Steps in Scale Development and Reporting: A Guide for Researchers', *Communication Methods and Measures*, vol. 12, no. 1, pp. 25-44.

Curzer, H. J. (2012) *Aristotle and the Virtues*, Oxford: Oxford University Press.

Darnell, C., Gulliford, L., Kristjánsson, K. and Paris, P. (2019) '*Phronesis* and the Knowledge-Action Gap in Moral Psychology and Moral Education: A New Synthesis?' *Human Development*, vol. 62, no. 3, pp. 101-129.

Darnell, C., Fowers, B. J. and Kristjánsson, K. (2022) 'A Multifunction Approach to Assessing Aristotelian *Phronesis* (Practical Wisdom)', *Personality and Individual Differences*, vol. 196, October, 111684.

De Caro, M., Marraffa, M., and Vaccarezza, M. S. (2021) 'The Priority of *Phronesis*: How to Rescue Virtue Theory from Its Critics', in M. S. Vaccarezza and De Caro, M. (eds.), *Practical Wisdom: Philosophical and Psychological Perspectives*, London: Routledge, pp. 29-51.

Fowers, B. J., Carroll, J. S., Leonhardt, N. D. and Cokelet, B. (2021) 'The Emerging Science of Virtue', *Perspectives on Psychological Science*, vol. 16, no. 1, pp. 118-147.

Grossmann, I. (2017) 'Wisdom in Context', *Perspectives on Psychological Science*, vol. 12, no. 2, pp. 233-257.

Harrison, T., Polizzi, G. and McLoughlin, S. (2023) 'Measuring Cyber-Wisdom: Preliminary Validation of a New Four-Component Measure', *Education and Information Technology*, https://doi.org/10.1007/s10639-023-11953-9

Jubilee Centre for Character and Virtues (2022). *A Framework for Character Education in Schools, 3rd ed.*, http://www.jubileecentre.ac.uk/userfiles/jubileecentre/pdf/character-education/Framework%20for%20Character%20Education.pdf

Kerr, S. (2021) 'Developing and Testing a Teaching Intermediate Concept Measure of Moral Functioning: A Preliminary Reliability and Validity Study', *Ethics and Behavior*, vol. 31, no. 5, pp. 350-364.

Krettenauer, T. (2019) 'The Gappiness of the "Gappiness Problem"', *Human Development*, vol. 62, no. 3, pp. 142-145.

Kristjánsson, K. (2020) *Flourishing as the Aim of Education: A Neo-Aristotelian view*, London: Routledge.

Kristjánsson, K. (2021) 'Twenty-Two Testable Hypotheses about *Phronesis*', *British Educational Research Journal*, vol. 47, no. 5, pp. 1303-1322.

Kristjánsson, K., Darnell, C., Flowers, B. J., Pollard, D. and Moller, F. (2020) *Phronesis: Developing a Conceptualisation and an Instrument. Research Report*, https://www.jubileecentre.ac.uk//userfiles/jubileecentre/pdf/Research%20Reports/Phronesis_Report.pdf

Kristjánsson, K., Pollard D., Darnell, C. and Thoma, S. (2021a) *Phronesis: Using an Aristotelian Model as a Research Tool. Research Report*, https://www.jubileecentre.ac.uk/userfiles/jubileecentre/Phronesis%20Report.pdf

Kristjánsson, K., Fowers, B., Darnell, C., and Pollard, D. (2021b) '*Phronesis* (Practical Wisdom) as a Type of Contextual Integrative Thinking', *Review of General Psychology*, vol. 25, no. 3, pp. 239-257.

Kristjánsson, K., Maile, A., Thompson, A. and McLoughlin, S. (2022) *Teaching Phronesis to Aspiring Police Officers. Research Report*, https://www.jubileecentre.ac.uk/userfiles/jubileecentre/pdf/Research%20Reports/23636%20-%20Jubilee%20Centre%20ASPIRING%20POLICE%20OFFICERS%20ST6_DIGITAL.pdf

Kristjánsson, K. and Fowers, B. J. (2024) *Phronesis: Retrieving Practical Wisdom in Psychology, Philosophy, and Education*, Oxford: Oxford University Press.

Miller, C. B. (2021) 'Flirting with Skepticism about Practical Wisdom', in M. S. Vaccarezza and De Caro, M. (eds.), *Practical Wisdom: Philosophical and Psychological Perspectives*, London: Routledge, pp. 52-69.

Moore, C., Detert, J. R., Klebe Treviño, L., Baker, V. L. and Mayer, D. M. (2012) 'Why Employees Do Bad Things: Moral Disengagement and Unethical Organizational Behavior, *Personnel Psychology*, vol. 65, no. 1, pp.1-48.

Paulhus, D. L., Buckels, E. E., Trapnell, P. D. and Jones, D. N. (2020) 'Screening for Dark Personalities', *European Journal of Psychological Assessment*, vol. 37, no. 3, pp. 208–222.

Morgan, B., Gulliford, L. and Kristjánsson, K. (2014) 'Gratitude in the U.K: A New Prototype Analysis and a Cross-Cultural Comparison', *Journal of Positive Psychology*, vol. 9, no. 4, pp. 281–294.

Peterson, A. (2020) 'Character Education, the Individual and the Political', *Journal of Moral Education*, vol. 49, no. 2, pp. 143-157.

Russell, D. C. (2009) *Practical Intelligence and the Virtues*, Oxford: Oxford University Press.

Sudbury-Riley, L. and Kohlbacher, F. (2016). 'Ethically Minded Consumer Behavior: Scale Review, Development, and Validation', *Journal of Business Research*, vol. 69, no. 8, pp. 2697-2710.

Tiberius, V. and Swartwood, J. (2011) 'Wisdom Revisited: A Case Study in Normative Theorizing', *Philosophical Explorations*, vol. 14, no. 3, pp. 277-295.

Wright, J., Warren, M. and Snow, N. (2021) *Understanding Virtue: Theory and Measurement*, Oxford: Oxford University Press.

Yan, V. X., Oyserman, D., Kiper, G. and Atari, M. (2023) 'Difficulty-as-Improvement: The Courage to Keep Going in the Face of Life's Difficulties', *Personality and Social Psychology Bulletin*, https://doi.org/10.1177/01461672231153680

> *PHRONESIS* IS A STATE GRASPING THE TRUTH, INVOLVING REASON, CONCERNING ACTION ABOUT WHAT IS GOOD OR BAD FOR A HUMAN BEING
>
> *Aristotle, Nicomachean Ethics, 1140b4-6*

Research Team

KRISTJÁN KRISTJÁNSSON
Professor and Principal Investigator, University of Birmingham

Kristján Kristjánsson is Professor of Character Education and Virtue Ethics, Jubilee Centre for Character and Virtues, University of Birmingham. He also holds fractional professorial positions at Boston College and University of Iceland. He is editor of the Journal of Moral Education.

STEPHEN THOMA
Professor Emeritus, University of Alabama, and Honorary Senior Research Fellow, University of Birmingham

Stephen Thoma is Professor Emeritus at the University of Alabama. He is an expert on neo-Kohlbergian theories of moral reasoning. He is also an Honorary Senior Research Fellow in the Jubilee Centre where he used to hold a professorial position.

SHANE MCLOUGHLIN
Assistant Professor, Jubilee Centre for Character and Virtues, University of Birmingham

Shane McLoughlin is an Assistant Professor at the Jubilee Centre for Character and Virtues, and a Chartered Psychologist. He specialises in psychometric measurement, assessment of intervention efficacy, meaning-focussed behaviour therapies, and individual differences. He is an Associate of the International Centre for Guidance Studies, and a Senior Fellow of Advance HE.

Acknowledgements

The report authors would like to thank the many people who have made the *Phronesis* Project possible, from the initial two reports to this final one.

Specifically, the authors would like to thank:

- Joseph McDowell, who designed this report.
- Jubilee Centre members of staff, past and present, who made a valuable contribution to the execution of the project and compilation of this report, in particular the Centre Director Professor James Arthur, Michael Fullard, Holly O'Donoghue, Danielle Edwards, Aidan Thompson, and Elliot Neiyyar.
- Professor Blaine J. Fowers who gave us some initial advice about the design of the project, insofar as it draws on the previous measure (LPM) to which he contributed.

JOHN TEMPLETON FOUNDATION
Inspiring Awe & Wonder

This project was made possible through the support of a grant from the John Templeton Foundation

For more information about the *Phronesis* project and the Jubilee Centre for Character and Virtues please contact:
tel: 0121 414 4875
email: jubileecentre@contacts.bham.ac.uk
www.jubileecentre.ac.uk

ISBN: 9780704429857

UNIVERSITY OF BIRMINGHAM

Jubilee Centre for Character and Virtues
12th Floor (West) Muirhead Tower
Edgbaston, Birmingham
B15 2TT, United Kingdom
www.birmingham.ac.uk

Designed and Formatted by
JMCDOWELL Engagement