The Four-Item Mentalising Index (FIMI) is a Valid, Reliable, and Practical Way to Assess Mentalising: Reply to Murphy et al. (2021)

Rachel A Clutterbuck[1], Lucy A Livingston[2], & Punit Shah[1*]

[1]Department of Psychology, University of Bath, Bath, UK.

[2]School of Psychology, University of Cardiff, Cardiff, UK.

*Corresponding author p.shah@bath.ac.uk

Department of Psychology,

University of Bath

Bath, BA2 4JD

Abstract

Murphy et al. (2022) raised concerns regarding the validity of the Four-Item Mentalising Index (FIMI). We wholeheartedly agree with Murphy et al. that there are problems in the social cognition literature hampering research. However, we maintain the FIMI is conceptually grounded in empirical and theoretical research, and that concerns regarding its construct validity are overstated. Drawing on recent research, we also discuss issues concerning the interpretation of discriminant validity analyses and discuss approaches to improve objectivity in future research. Overall, we argue that the FIMI is a conceptually and psychometrically sound measure, which has utility in future social cognition research and clinical practice.

Public significance statement

We discuss a recent critique of a new mentalising measure, the Four-Item Mentalising Index (FIMI). We agree there are issues within the social-cognition literature, but argue that the criticisms put forth about the FIMI are overstated. Using recent literature to support our arguments, we maintain that the FIMI is a valid, reliable, and valuable new tool for measuring mentalising.

We developed and validated the first self-report measure of mentalising ability, the Four-Item Mentalising Index (FIMI; Clutterbuck et al., 2021). Concerns regarding the conceptual underpinnings and validity of the FIMI were raised by Murphy et al. (2022), stimulating broader discussion about measuring social cognition. Here, we draw upon on Murphy et al.'s paper to reiterate the utility of the FIMI, as well as advance discussions about social cognition theory and measurement in light of the recent research.

Murphy et al. argue that inconsistent use of terminology and conceptualisations of social-cognitive constructs have created challenges in social cognition research. We agree with this argument, and the inconsistent conceptualisation of whether mentalising encompasses the understanding of affective (emotional) states as well as cognitive (non-emotional) states is a central problem in this literature. Murphy et al. claim that our conceptualisations of mentalising and cognitive empathy are too narrow, ignoring overlap between the affective and cognitive elements of these constructs. Yes, mentalising was traditionally thought to have a cognitive and an affective component, which underpins Murphy et al.'s thesis. Critically, however, newer research suggests that the affective component of mentalising is conceptually synonymous with cognitive empathy (e.g., Preston et al., 2020; Schurz et al., 2021). A large body of research now supports a distinction between [cognitive] mentalising and cognitive empathy at the conceptual (e.g., Reniers et al., 2011), biological (e.g., Sebastian et al., 2012; Shamay-Tsoory et al., 2010), and behavioural (e.g., Torralva et al., 2015) levels. Preston et al. (2020), for example, argue that mentalising tasks which have emotional content are in fact measuring cognitive empathy. Modelling the relationship between mentalising and cognitive empathy, several authors have suggested that mentalising is a prerequisite for cognitive empathy (e.g., Sebastian et al., 2012; Shamay-Tsoory et al., 2010), supporting the notion that mentalising may be measured independently of cognitive empathy, but not the other way around.

Murphy et al. also argue that a distinction between mentalising and cognitive empathy is unlikely to be made through self-report measures. However, self-report measures can allow for the careful regulation of emotional content, and therefore a more controlled investigation of these separable constructs. As the first validated self-report mentalising measure, the FIMI provides an opportunity to investigate research questions such as these, i.e., quantifying common and unique variance between self-reported mentalising and empathy. Further, our item selection method, which employed the judgements of four expert raters to identify items without emotional content, were crucial to ensure the validity of the FIMI with respect to the mentalising construct. Murphy et al.'s claim that we used "eyeball judgements" to select items grossly misrepresents the methodological rigour and quantitative elements of this approach (e.g., high inter-rater reliability). Indeed, we propose that such an approach, not commonly used in the field of social cognition research, might be useful in future research when developing self-report measures. In a similar manner, Murphy et al. also understate the importance of the statistical methods used to psychometrically validate the FIMI, including its reliability. Such analyses are often not performed as thoroughly in the field of social cognition research, and offer significant advances for the validation of novel social cognition measures. Hence, we maintain that the FIMI represents an important, not redundant, advance to psychological assessment in this field.

In contrast to this approach, behavioural measures of mentalising, which Murphy et al. imply are superior to self-report measures, often conflate emotional and non-emotional content. For example, the Movie for the Assessment of Social Cognition (MASC; Dziobek et al., 2006) quantifies emotional and non-emotional social cognition scores, yet it does not hold constant emotional and non-emotional content. While social-cognitive self-report measures have the same limitations as all self-report measures, they also have practical advantages, and their flaws are arguably more constrained and transparent than problems in behavioural

measures that can be harder to detect (e.g., putative social-cognitive tasks measuring domain-general cognitive mechanisms; Heyes, 2014). Self-report measures are also better able to capture individual differences and measure *typical* rather than *optimal* cognitive function (Dang et al., 2020). In developing a self-report measure of mentalising, we avoided adding to the "arms race" (see Gernsbacher & Yergeau, 2019, p. 109) of behavioural mentalising measures that are overly complex, unsuitable for online data collection, and often measuring lower-level cognitive processes (e.g., Heyes, 2014; Quesque & Rosetti, 2020). The primary motivation for developing the FIMI was so that it can be used to collect very large samples of online data, without sacrificing the range of individual variability captured by the scale, which has been difficult to achieve with existing measures. Together, in the face of this important contribution, Murphy et al.'s (2022) claims that the FIMI adds nothing to existing social cognition measures are overstated.

Murphy et al. also raise other concerns regarding the construct validity of the FIMI. They suggest that findings showing that autistic people score significantly lower on the FIMI than non-autistic people do not support the measure's construct validity, as lower scores in autistic people would also be expected in relation to empathy. Surprisingly, however, they neglect the literature showing that empathy is increasingly considered to be intact in autism, despite ongoing mentalising atypicalities (see Livingston & Happé, 2021). Indeed, there is a large and burgeoning literature suggesting that mentalising difficulties in autism are more pronounced than empathetic difficulties (e.g., Brewer et al., 2015), which effectively speaks against Murphy et al.'s concern about the construct validity of the FIMI.

Similarly, Murphy et al. raise concerns about the validity of using the refined RMET (Olderbak et al., 2015) to demonstrate convergent validity with the FIMI, given that this measure has an emotional component. We recognise Murphy et al.'s concerns regarding the emotional content of the RMET stimuli. It is to be emphasised that the refined RMET has

5

acceptable psychometric properties, mitigating the most serious concerns associated with the original RMET (e.g., Olderbak et al., 2015). It is also particularly suitable for online data collection and there are few behavioural measures of mentalising and social cognition that have been psychometrically validated as well as this recent measure. Nonetheless, the concerns raised by Murphy et al. regarding the abbreviated RMET further highlight the need for better online mentalising measures, which would enable us to collect larger and more representative samples of data, as well as access clinical participants who may not wish to participate in-person (see also, Livingston et al., 2019). Since the publication of the FIMI, an online version of the Frith-Happé mentalising task has been developed (Livingston et al., 2021), which shows promise to advance online data collection using behavioural measures. However, there is a long way to go in addressing challenges in the field given the well-documented issues with behavioural mentalising measures correlating weakly with each other (e.g., Warnell & Redcay, 2019) and showing poor construct validity (e.g., Heyes, 2014). Until these issues are addressed, establishing the validity of novel mentalising measures remains a serious challenge as there is a paucity of theoretically grounded tasks to validate new measures against. Overcoming this issue is not straightforward, especially given that the structure of social cognition is not yet fully understood (e.g., Schurz et al., 2021). To this end, given its conceptual and face validity, we expect that the FIMI will contribute to the refinement of existing or development of novel behavioural measures of mentalising.

Murphy et al. argue that correlations between the FIMI and other cognitive empathy measures are too high and speak against the FIMI's discriminant validity. Crucially, however, assessing discriminant validity requires close consideration of theoretical underpinnings, common method variance, and other contextual factors (Clark & Watson, 2019; Rönkkö & Cho, 2022). Notably, *"a moderately small correlation between measures does not always imply that two constructs are distinct, and a high correlation does not imply that they are*

not" (Rönkkö & Cho, 2022, p33). In line with this recent literature (Clark & Watson, 2019; Rönkkö & Cho, 2022), we argue that the theoretical grounding of our mentalising construct provides a clear distinction from cognitive empathy to assuage fundamental concerns about the FIMI's construct validity. Further, the strength of the correlations between the FIMI and cognitive empathy reported by Murphy et al. are likely to be a result of i) similar methods of measurement used between variables (questionnaire measures using similar or identical Likert scales), ii) an expected and acceptable degree of overlap between the constructs, and iii) duplicate data between variables in Murphy et al.'s dataset which may have inflated the correlations, leading to a misrepresentation of the strength of the relationship between the FIMI and cognitive empathy. Similarly, Murphy et al. argue that, because the FIMI items do not form a latent factor within the cognitive empathy subscale of the measure from which they were adapted, the FIMI is simply redundant to the measure of cognitive empathy. However, a clear-cut distinction between the FIMI's items and the remaining items from the measure of cognitive empathy was never expected through data driven discriminant validity analyses (as we noted in Clutterbuck et al., 2021), due to the shared method variance and conceptual overlap. Rather, the FIMI items were necessarily selected through a close and theoretically informed examination by expert raters, to reflect the clear-cut conceptual distinction between mentalising and cognitive empathy. Taken together, we argue that the FIMI should not be considered redundant with the cognitive empathy measures used by Murphy et al. Rather, we suggest that Murphy et al.'s results be interpreted in the appropriate theoretical and methodological context.

On a final note, Murphy et al. claim that similarity in the correlations between cognitive and affective empathy, with the correlations between the FIMI and affective empathy, show that the FIMI does not measure a distinct construct from cognitive empathy. This appears to be based on the assumption that cognitive empathy is conceptually closer to

affective empathy than mentalising. However, Murphy et al. do not cite compelling evidence to support this assumption. And empirical findings, in non-clinical samples, do not show stronger correlations between cognitive and affective empathy, when compared to mentalising and affective empathy (e.g., Shamay-Tsoory et al., 2009), speaking against Murphy et al.'s claims. Similarly, Murphy et al. argue that significant correlations between the FIMI and personality traits in their data speak against the measure's construct validity, since, they claim, such relationships are not often found with behavioural measures of mentalising. This argument, like others put forward by Murphy et al., shows a selective reading of the existing literature. Behavioural measures of mentalising *have* shown similar relationships with personality traits (e.g., Vonk et al., 2015), including agreeableness (Nettle & Liddle, 2008), and there are several potential explanations for this. Stronger mentalising ability may enable individuals to be more agreeable through having better social skills. Alternatively, people who are more agreeable may be more motivated to attend to the mental states of others, thereby enhancing their mentalising skills (Nettle & Liddle, 2008). It would be of interest to elucidate the nature of these associations in both future research and clinical practice which, again, the FIMI should help rather than hinder.

Overall, we thank Murphy et al. for reiterating some important issues within the field of social cognition, including the inconsistent use of terminology and the various, sometimes contradictory, models of social cognition. These issues, together with recognised inconsistencies in the construct validity of social-cognitive measures (e.g., Quesque & Rosetti, 2020; Warnell & Redcay, 2019), are a major problem in the field, discussions of which are already ongoing elsewhere in the literature (e.g., Happé et al., 2017). It is clear that the structure of social cognition is not yet fully understood, but research continues to advance our understanding of the in(ter)dependence of social-cognitive constructs (e.g., Schurz et al., 2021). Therefore, as the first self-report measure of social cognition to separate mentalising

from empathy based on emotional content, the FIMI offers an important contribution to the measurement of social cognition. The FIMI is especially designed to improve understanding of the relationships between affective and cognitive socially relevant processes in people with clinical conditions, like autism.

In future, Murphy et al. argue for discriminant validity tests to be a "non-negotiable requirement" of test development. However, for this to happen, there first needs to be closer consideration of how to interpret discriminant validity analyses in the context of theoretical and measurement similarity between constructs. As Murphy et al. note themselves, whether a correlation coefficient is too high or too low is generally up to the interpretation of the researcher, and it is critical that methodological and conceptual overlaps are considered in this interpretation. Evaluating conceptual similarity between constructs is particularly important for social-cognitive abilities because these are frequently found to relate to various social and psychological factors; for example, personality (e.g., Vonk et al., 2015), socio-demographics (Dietze & Knowles, 2021), and political beliefs (Hasson et al., 2018). To maximise objectivity in interpretations of convergent and discriminant validity, we suggest that researchers more carefully focus on overlaps between social-cognitive constructs at several levels of explanation, such that discriminant validity analyses can be interpreted in their appropriate context (see also, Happé et al., 2017).

Altogether, we agree with Murphy et al. that there are fundamental problems within the social cognition literature, however the FIMI will do more to overcome than exacerbate this problem. Our careful choice of conceptual terminology, suitably drawing on theoretical frameworks across multiple levels of explanation (e.g., biological to behavioural), and the use of robust statistical methods are firmly grounded in recent developments in psychological assessment. We therefore argue that the FIMI will make a unique and valuable contribution to both psychological research and clinical practice, given its i) close conceptual alignment

with current theories on mentalising and social cognition, ii) robust psychometric properties,

and iii) practical utility in online or time-restricted settings.

References

Brewer, R., Happé, F., Cook, R., & Bird, G. (2015). Commentary on "Autism, oxytocin and interoception": Alexithymia, not Autism Spectrum Disorders, is the consequence of interoceptive failure. *Neuroscience & Biobehavioral Reviews*, *56*, 348-353.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, *31*(12), 1412.

Clutterbuck, R. A., Callan, M. J., Taylor, E. C., Livingston, L. A., & Shah, P. (2021). Development and validation of the Four-Item Mentalising Index. *Psychological Assessment, 33*(7), 629–636.

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated?. *Trends in Cognitive Sciences*, *24*(4), 267-269.

Dietze, P., & Knowles, E. D. (2021). Social class predicts emotion perception and perspective-taking performance in adults. *Personality and Social Psychology Bulletin*, *47*(1), 42-56.

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., Kessler, J., Woike, J. K., Wolf, O. T., & Convit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623-636.

Gernsbacher, M. A., & Yergeau, M. (2019). Empirical failures of the claim that autistic people lack a theory of mind. *Archives of Scientific Psychology, 7*(1), 102-118.

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In(ter) dependence of sociocognitive processes. *Annual Review of Psychology*, *68*, 243-267.

Hasson, Y., Tamir, M., Brahms, K. S., Cohrs, J. C., & Halperin, E. (2018). Are liberals and conservatives equally motivated to feel empathy toward others? *Personality and Social Psychology Bulletin*, *44*(10), 1449-1459.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science*, *9*(2), 131-143.

Livingston, L. A., Carr, B., & Shah, P. (2019). Recent advances and new directions in measuring theory of mind in autistic adults. *Journal of Autism and Developmental Disorders*, *49*(4), 1738-1744.

Livingston, L. A., & Happé, F. (2021). Understanding atypical social behaviour through social cognitive theory: Lessons from autism. In H. J. Ferguson, V. E. A. Brunsdon, & E. E. F. Bradford (Eds.), *The cognitive basis of social interaction across the lifespan.* Oxford University Press.

Livingston, L. A., Shah, P., White, S., & Happé, F. (2021). Further developing the Frith-Happé animations: A quicker, more objective, and web-based test of theory of mind for autistic and neurotypical adults. *Autism Research*, *14*(9), 1905-1912.

Murphy, B. A., Hall, J. A., & Duong, F. (2022) It Looks Like Construct Validity, But Look Again: Comment on Clutterbuck et al. (2021) and Recommendations for Test Developers in the Broad "Empathy" Domain. *Psychological Assessment*.

Nettle, D., & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality*, *22*(4), 323-335.

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. *Frontiers in Psychology*, *6*, 1503.

Preston, S. D., Ermler, M., Lei, Y., & Bickel, L. (2020). Understanding empathy and its disorders through a focus on the neural mechanism. *Cortex*, *127*, 347-370.

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science*, *15*(2), 384-396.

Reniers, R. L., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84-95.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6-47.

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J., & Kanske, P. (2021). Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin, 147*(3), 293-327.

Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S. J., De Brito, S. A., McCrory, E. J., & Viding, E. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*(1), 53-63.

Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J., & Levkovitz, Y. (2010). The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex*, *46*(5), 668-677.

Shamay-Tsoory, S., Harari, H., Szepsenwol, O., & Levkovitz, Y. (2009). Neuropsychological evidence of impaired cognitive empathy in euthymic bipolar disorder. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *21*(1), 59-67.

Torralva, T., Gleichgerrcht, E., Ardila, M. J. T., Roca, M., & Manes, F. F. (2015). Differential cognitive and affective theory of mind abilities at mild and moderate stages of behavioral variant frontotemporal dementia. *Cognitive and Behavioral Neurology*, *28*(2), 63-70.

Vonk, J., Zeigler-Hill, V., Ewing, D., Mercer, S., & Noser, A. E. (2015). Mindreading in the dark: Dark personality features and theory of mind. *Personality and Individual Differences*, *87*, 50-54.

Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition*, *191*, 103997.