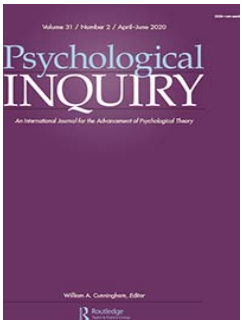

Routledge
Taylor & Francis Group

Psychological Inquiry

An International Journal for the Advancement of Psychological Theory

ISSN: 1047-840X (Print) 1532-7965 (Online) Journal homepage: https://www.tandfonline.com/loi/hpli20

"A Theory of Wisdom Needs Theory of Mind"

Jeffrey Lees & Liane Young

To cite this article: Jeffrey Lees & Liane Young (2020) "A Theory of Wisdom Needs Theory of Mind", Psychological Inquiry, 31:2, 168-173, DOI: 10.1080/1047840X.2020.1750925

To link to this article: https://doi.org/10.1080/1047840X.2020.1750925

Published online: 22 Jun 2020.

PSYCHOLOGICAL INQUIRY
2020, VOL. 31, NO. 2, 168–173
https://doi.org/10.1080/1047840X.2020.1750925

Routledge
Taylor & Francis Group

"A Theory of Wisdom Needs Theory of Mind"

Jeffrey Lees[a] and Liane Young[b]

[a]Harvard Business School, Harvard University, Cambridge, Massachusetts; [b]Department of Psychology, Boston College, Boston, Massachusetts

Introduction

The Common Wisdom Model (Grossmann et al., this issue) represents a significant synthesis and advancement of our scientific understanding of wisdom. By arguing that moral aspirations and perspectival meta-cognition (PMC) are the foundational psychological components of wisdom, the Common Wisdom Model presents a roadmap for scholars who hope to study wisdom, its antecedents and correlates, and examine how wisdom can be promoted and maintained through scientific and educational interventions. Moreover, the Common Wisdom Model opens the doors to scholars in separate research traditions, such as those who study moral psychology or meta-cognition in isolation, thereby contributing to the growing interdisciplinary community of scientists interested in understanding wisdom as a psychological phenomenon.

Of course, our task at hand is to present a more critical analysis, and so this commentary focuses not on our alignment with the model put forth in the target article but by and large our concerns. Our primary concerns regard the model's approach to defining moral aspirations and PMC, and the lack of attention paid to their relationship. We argue that PMC is too narrowly defined, insofar as it excludes theory of mind ("ToM"), and that moral aspirations require a more explicit definition. These issues become paramount in the context of the underspecified relationship between PMC and moral aspirations. Moral cognition is intimately related to ToM—some would say ToM is "essential" (Gray, Young, & Waytz, 2012; Schein & Gray, 2018)–and therefore we argue that excluding ToM from the definition of PMC is problematic and prevents the Common Wisdom Model from fully explicating the morality-PMC relationship. If the study of wisdom is to reveal how wisdom is fostered and maintained, the relationship between the components of the Common Wisdom Model must be better understood, and such understanding requires integrating ToM into the conception of PMC.

Clarifying the Components of the Common Wisdom Model

What is the Moral Aspirations—PMC Relationship?

Perhaps one of the most important questions for the Common Wisdom Model that is left mostly unanswered concerns the nature of the relationship between moral aspirations and PMC, the two foundational components of the model. Understanding this relationship would seem essential for answering the scientific question of how wisdom develops. How can we hope to foster wisdom through scientific intervention if we do not know how, or even whether, the two central psychological components of wisdom interact?

At first, moral aspirations and PMC are presented as separate, orthogonal components of the model, suggesting that we should imagine some individuals may be high in one but low in the other. If this is the case, perhaps those high in moral aspirations but low in PMC are individuals epitomized by the aphorism "the road to hell is paved with good intentions," and conversely those low in moral aspirations but high in PMC are Machiavellian-like social manipulators. This conception would also suggest that efforts aimed at fostering wisdom must focus simultaneously on building moral aspirations and PMC by targeting the (putatively) different mechanisms underlying their respective psychological development.

Yet at multiple points it is suggested that PMC and moral aspirations are causally linked, and often in contradictory directions. Moral aspirations are often presented in the Common Wisdom Model as motivational drivers of PMC, with PMC being oft-described as "morally-grounded," suggesting that without moral aspirations individuals are unlikely to engage in PMC. But at other times it is argued that "PMC is required to implement wisdom-related moral aspiration" and that "PMC can override the immediate impulse to protect self-interests," suggesting that PMC is a necessary precondition for wisdom-like moral aspiration to arise. It is difficult to surmise how the Common Wisdom Model conceptualizes the moral aspirations–PMC relationship because the mere question of their relationship is not explicitly addressed, nor are further questions of directionality, orthogonality, and/or causality regarding the two central components of the model.

The lack of clarity regarding the moral aspirations–PMC relationship is compounded by the vague definition provided for moral aspirations alongside the narrow definition provided for PMC. Without specific psychological detail, it is difficult to discuss moral aspirations either on their own or in relation to other psychological processes. And by intentionally excluding other-focused meta-cognition (i.e., theory of mind) from the formal definition of PMC, the

CONTACT Jeffrey Lees jlees@g.harvard.edu Wyss House, Harvard Business School, Boston, MA 02163, USA.

Common Wisdom Model neglects an abundant and relevant body of research on the relationship between meta-cognition and moral cognition.

Meta-Cognition Without Theory of Mind?

The authors of the Common Wisdom Model state, at multiple points, that their definition of perspectival meta-cognition (PMC) refers exclusively to meta-cognition directed toward *the self*, not toward others. In the model, PMC is a dialogical, self-reflective process that leads to the consideration of multiple, divergent perspectives related to the situation at hand, while also engendering a level of epistemic/intellectual humility. The distinction between self-focused and other-focused meta-cognition has long been recognized (Kitchener, 1983), and the authors point to some preliminary evidence that "perspective-taking" in the wisdom literature (i.e., PMC) is a distinct process from "perspective-taking" in the theory of mind literature (i.e., other-focused meta-cognition) (Rakoczy, Wandt, Thomas, Nowak, & Kunzmann, 2018). While we see the value in making this intellectual distinction, in the context of the Common Wisdom Model we argue that ToM *should not* be excluded from the definition of PMC, as doing so impedes the model's ability to consider the relationship between meta-cognition and moral aspirations as they relate to wisdom.

For the sake of clarity, here we use the term "theory of mind" (ToM) to refer to a broad range of (primarily) other-focused meta-cognitive processes studied across the extant psychological literature, including mind perception (Epley & Waytz, 2010), cognitive and affective empathy (perspective-taking/empathic concern: Davis, 1983; Decety, 2011; Epley, Keysar, Van Boven, & Gilovich, 2004), meta-perception (Malloy, Albright, Kenny, Agatstein, & Winquist, 1997; Vazire & Carlson, 2011), and ToM as a developmental phenomenon in children (Wellman, Cross, & Watson, 2001). It is worth noting, however, that ToM research has primarily studied other-focused meta-cognition more as a matter of practice than as an in-principle psychologically distinct process from self-focused meta-cognition (For discussion of how ToM processes may apply to self versus other, see Gweon, Young, and Saxe (2011) and Young and Tsoi (2013)).

Despite this conceptual distinction between PMC and ToM, and the Common Wisdom Model's insistence that ToM is not a noteworthy component of PMC or wisdom more broadly, the authors inadvertently invoke ToM processes at multiple points throughout the paper. For example, the authors cite the classic false-consensus effect (Ross, Greene, & House, 1977) as an early example of the type of cognitive bias that wisdom would help overcome. Yet the false-consensus effect does not represent a mere egocentric bias (i.e., poor PMC); rather, it is an egocentric bias in ToM judgments of others' preferences. Elsewhere, the authors argue that wisdom may have developed because of evolutionary pressures to "consider different viewpoints and to reconcile them with one's own viewpoints". Yet, the question of how someone is able to consider "different viewpoints," a cognitive ability that is repeatedly presented as a defining feature of PMC and wisdom more broadly, without engaging in ToM is not clear. And even if we accepted that ToM is implied *ipso facto* in the idea of "considering different viewpoints" the authors go out of their way to separate ToM from PMC, i.e., the terminology "theory of mind" appears three times in the target article, and each is an instance where the authors argue it is theoretically and empirically separate from PMC.

Instances in which the authors hew to their more narrow definition of PMC though also illustrate its further problems. Early in the paper the authors present a hypothetical moral dilemma where an individual needs to board a train in order to deliver wedding rings in time for the wedding ceremony but has no way of obtaining a train ticket other than to steal a ticket from an individual who (a) is distracted, and (b) looks sufficiently wealthy as to not experience the theft as financially burdensome. The authors claim that an unwise person would simply apply a "stealing is wrong" logic to this dilemma, whereas a wiser person engaging in PMC would also consider the moral value of loyalty, in that *not* stealing the ticket in order to arrive with the wedding rings on time could itself be considered a violation of a moral obligation to be loyal to one's friends.

It is hard to see how this example is meta-cognitive, whether using the narrow definition of PMC provided by the authors or a more expansive definition of meta-cognition. Instead, the example seems to merely involve the weighing of multiple decision alternatives. Someone patronizing a fast-food restaurant who is considering competing utilities of deliciousness versus health in their food choice is certainly engaging in more nuanced and dialogical thought, but this process does not necessarily involve thinking about thought (meta-cognition). Even if we accept that considering two competing moral values (rightness of loyalty vs. wrongness of theft) in the train-ticket-wedding-rings example is an instance of PMC as defined by the authors, we are left asking how the application of such PMC led the wiser individual to recognize loyalty as a situationally relevant moral value, and answering this question *without* invoking ToM is tricky. It cannot be the case that the wise individual prospectively realized that their friend might perceive the decision not to steal as disloyal, since that would be an act of ToM—moral meta-perception specifically (Rom, Weiss, & Conway, 2017). It cannot be the case that the wise individual considered how a hypothetical third-party would behave in this situation, or how the person they would be stealing from might feel, since those are acts of ToM. It cannot be the case that the wise individual is recalling a past moral dilemma where they did not consider loyalty and others reacted negatively, because that would have required accurate moral meta-perception at the time. And it cannot be the case that the wise individual is recalling the negative affect they experienced from a past instance where they were the victim of disloyalty, as this would require realizing that others might too feel this negative affect, which is an act of ToM.

The only way to explain how this wise individual realized loyalty was a situationally relevant moral value without invoking ToM is if the wise individual learned about loyalty from a book; they were the subject of some pedagogical intervention that taught them what loyalty was, how to identify when it is relevant in any situation, and convinced them it was a value worth pursuing. While this is not impossible, at face value, we think it is much less likely than the ToM-driven explanations above, and in fact resembles the propositional logic ("stealing is wrong") for which the authors consider this whole moral dilemma a counterexample.

The purpose of the exercise above is not to be definitionally pedantic but rather to demonstrate that excluding ToM from the Common Wisdom Model and PMC, specifically, leaves a paucity of explanations for something as simple as considering two discrete moral values. By integrating ToM into the definition PMC we can begin to consider the nature of the meta-cognition–moral aspirations relationship as informing wisdom, and in doing so draw from the large body of scholarship explicating how ToM drives moral cognition.

Theory of Mind and Morality Are Intimately Linked and Related to Wisdom

Moral psychologists have long noted the intimate connection between theory of mind and moral cognition (for a recent review see Kim, Park, & Young, 2020). One tradition in moral psychology, dyadic morality theory, argues that mind perception (mental state attributions toward others, i.e., ToM) is the "essence of morality" and constitutes the fundamental cognitive building blocks of moral judgment (Gray et al., 2012; Schein & Gray, 2018). At a conceptual level we argue it is easy to see how ToM is central to many components of morality. Judging whether an action caused harm is an act of ToM (Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014; Schein & Gray, 2016), as are more explicit judgments of others' moral intent (Ames & Fiske, 2015; Chakroff et al., 2016) and motives (Reeder, Vonk, Ronk, Ham, & Lawrence, 2004; Waytz, Young, & Ginges, 2014). In order to imagine moral judgment proceeding without consideration of the thoughts and feelings of others, one might imagine a strict deontologist who considers *only* the application of abstract, propositional moral rules. Yet even with deontological moral decision-making we know that ToM plays a role, both through the automatic cognitive association between "wrong" and "harmful to others" (Gray & Schein, 2012) and through reputational concerns (Lee, Sul, & Kim, 2018).

While the evidence for the relationship between ToM and morality is overwhelming, just as important for the Common Wisdom Model is evidence that ToM plays a role in the relationship between moral behavior and wisdom/wisdom-related constructs. For example, intellectual humility is cited as a core component of PMC and wisdom more broadly, and a core component of intellectual humility is respect for others' viewpoints (Krumrei-Mancuso & Rouse, 2016), which in turn requires both ToM and moral aspirations. Moreover, the relationship between intellectual humility and prosocial behavior is mediated by both cognitive and emotional empathy (Krumrei-Mancuso, 2017), but not explained by a multitude of other psychological traits and factors (Exline & Hill, 2012). Additionally, several scale measures of intellectual humility and wisdom find cognitive-empathy (trait perspective-taking) to be one of the largest correlates (Brienza, Kung, Santos, Bobocel, & Grossmann, 2018; LaBouff, Rowatt, Johnson, Tsang, & Willerton, 2012). These findings in the intellectual humility literature strongly suggest that the role of ToM processes in the psychology of wisdom is unexplored by the Common Wisdom Model.

In summary, while we see the intellectual value in distinguishing self-focused meta-cognition (PMC) from other-focused meta-cognition (typically studied as ToM) we argue that excluding ToM from PMC and the Common Wisdom Model does not make empirical or theoretical sense and serves only to obfuscate, rather than illuminate, the relationship between the Common Wisdom Model's two central components: PMC and moral aspirations. Below we discuss what integrating ToM into the Common Wisdom Model might look like, how this approach might improve the model, and future research directions for wisdom scholars seeking to understand the ToM—Wisdom relationship.

Centering Theory of Mind in the Common Wisdom Model

Wisdom and Accuracy in Theory of Mind Judgments

Integrating ToM into the Common Wisdom Model, and specifically the conception of PMC, requires first asking the question of what dimensions of ToM are expected to covary with wisdom. In the Common Wisdom Model, PMC covaries with wisdom at the level of activation, in that the unwise are less likely to engage in PMC than the wise, although it is unclear the extent to which this represents the motivation or the capacity to engage in PMC. However, unlike PMC, ToM varies less at the level of activation than it does at the level of *accuracy,* as it is subject to a host of well-documented biases. Taking others' perspectives can be a real cognitive challenge (Eyal, Steffel, & Epley, 2018), including when trying to understand how others are perceiving one's own self (Vazire & Carlson, 2011). ToM is subject to a host of systematic biases including self-anchoring/projection (Ames, 2004), negativity biases (Lees & Cikara, 2020), reliance on stereotypes (Lewis, Hodges, Laurent, Srivastava, & Biancarosa, 2012) and normative expectations (Furr, 2008), along with additional attribution and actor-observer asymmetries (Malle, Knobe, & Nelson, 2007; Waytz et al., 2014).

Integrating ToM into the conception of PMC leads to the reasonable theoretical conclusion that those who are wise are *more accurate* in judging the thoughts and feelings of others. To our knowledge there is no empirical research that directly examines the relationship between ToM judgment accuracy and existing measures of wisdom. In our opinion, this represents a fruitful and vital area of future research.

Such research is not merely for its own sake but represents a direct test, we think, of the Common Wisdom Model. The authors assert that "wisdom has been associated with deep interests in knowing the truth and with compassionate concerns for others' welfare," a statement that nicely dovetails with a theoretical ToM accuracy–wisdom relationship, and its function in connecting meta-cognition to moral aspirations. The Common Wisdom Model consistently proposes that wisdom is about considering others' perspectives in specific contexts, which would seem difficult without robust abilities in social perception and ToM. The authors also argue that PMC is better conceptualized as a state rather than as a trait construct, which parallels research on ToM accuracy suggesting that accuracy is less a function of domain-general capacity and more a function of the social relationship between the observer and the individual being judged (Carlson, 2016; Eyal et al., 2018; Zaki, Bolger, & Ochsner, 2008). Lastly, self-knowledge is also highlighted in the Common Wisdom Model, and research on self-knowledge suggests that meta-perception, critically involving ToM, plays a central role in how we come to understand ourselves (Vazire & Carlson, 2010).

We wish to stress again just how much research there is to be done in examining the relationship between judgment accuracy (including ToM accuracy) and wisdom. Accuracy in judging others is one of the oldest research traditions in social psychology (e.g., Cronbach, 1955), and continues to flourish today (e.g., Brashier & Marsh, 2020). There are numerous methodological and theoretical perspectives to be leveraged by wisdom researchers in examining how wisdom relates to judgment accuracy, broadly defined. We suggest reviewing West and Kenny (2011), Biesanz (2010), and Barranti, Carlson, and Côté (2017) for up-to-date methodological approaches to measuring judgment accuracy that can be straightforwardly and productively integrated into research on wisdom.

Connecting Morality and Wisdom via Theory of Mind

Integrating ToM into the Common Wisdom Model's conception of PMC will also contribute to an understanding of the unaddressed relationship between moral aspirations and PMC. Scholars continue to debate the directional relationship between ToM and moral judgment (e.g., Phillips & Knobe, 2018), but for the sake of illustration, we will adopt the position of dyadic morality theory: ToM *causes* moral judgment. If we accept the theoretical view that PMC either (a) leads moral aspirations to develop, or more conservatively (b) allows for the synthesis and application of existing moral aspirations, this has clear implications for the study of wisdom.

One of the most immediate implications relates to fostering wisdom. If meta-cognition is a necessary antecedent to moral aspirations then efforts to foster wisdom should focus primarily on developing PMC. The value of such an approach has already been suggested by the evidence that inducing PMC increases cooperation (Grossmann, Brienza, & Bobocel, 2017). This directional relationship also allows us to reinterpret previous findings. In one study (Bruya & Ardelt, 2018), students who completed a self-reflection journal on understanding the values and beliefs of exemplars of wisdom scored higher on measures of wisdom than did students who completed self-cultivation journals focused on character strengths. The authors of the Common Wisdom Model cite this finding as evidence that self-reflection induced wisdom, but we note that this intervention also involved ToM. Specifically, the wisdom-inducing intervention asked students to consider the perspectives of specific others who exemplify wisdom, which is an act of ToM. Whether or not ToM accounts for these specific findings, we urge wisdom researchers to consider the role that ToM plays in wisdom, and the directionality of the PMC–moral aspirations relationship, when developing interventions designed to foster and maintain wisdom.

Even given a more agnostic view with respect to the directional relationship between meta-cognition and moral cognition, integrating ToM into the Common Wisdom Model will clarify many of the intermediary processes described by the model. Including ToM in the Common Wisdom Model gives us a roadmap for understanding many critical details that the model is currently unable to provide, such as what specific mental state attributions are the wise making toward situational others (e.g., how are the wise interpreting the behavior of others and forecasting their future behavior), what specific moral aspirations are the wise applying to the situation at hand (e.g., how are the wise judging the potential harms and benefits of their behavior for others), what specific perspectives are the wise choosing to consider and able to perceive (e.g., how are the wise determining that specific others have perspectives different from their own), and how do these specific processes differ from those of the unwise. These types of questions require answers if the Common Wisdom Model is to provide a discerning and predictive framework for guiding wisdom research over the coming decade, and we believe that accommodating ToM will be essential to this challenge.

Conclusion

Despite our critique, we are genuinely enthusiastic about the integrity of the Common Wisdom Model and its ability to guide important future work on wisdom. If anything, we feel the model does not go *far enough* in asserting that morality and meta-cognition are the foundations of wisdom! Perhaps counterintuitively, we believe the model can and should capture a broader turf, beyond its narrow definition of perspectival meta-cognition. We argue that by expanding the model's definition of meta-cognition to include theory of mind processes the model will be in an even more powerful position to articulate the relationship between moral aspirations and meta-cognition, to make specific predictions regarding the cognitive processes underlying wisdom, and to pave the foundation for developing interventions to foster and maintain wisdom. We are optimistic about the success of the Common Wisdom Model, and we look forward to

the following decade's advancement in the science of wisdom.

Acknowledgements

We are thankful to Kevin Jiang for helpful comments on earlier drafts.

References

Ames, D. R. (2004). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, *87*(5), 573–585. doi:10.1037/0022-3514.87.5.573

Ames, D., & Fiske, S. (2015). Perceived intent motivates people to magnify observed harms. *Proceedings of the National Academy of Sciences*, *112*(12), 3599–3605. doi:10.1073/pnas.1501592112

Barranti, M., Carlson, E. N., & Côté, S. (2017). How to test questions about similarity in personality and social psychology research: Description and empirical demonstration of response surface analysis. *Social Psychological and Personality Science*, *8*(4), 465–475. doi:10.1177/1948550617698204

Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, *45*(5), 853–885. doi:10.1080/00273171.2010.519262

Brashier, N. M., & Marsh, E. J. (2020). Judging truth. *Annual Review of Psychology*, *71*(1), 499–515. doi:10.1146/annurev-psych-010419-050807

Brienza, J. P., Kung, F. Y. H., Santos, H. C., Bobocel, D. R., & Grossmann, I. (2018). Wisdom, bias, and balance: Toward a process-sensitive measurement of wisdom-related cognition. *Journal of Personality and Social Psychology*, *115*(6), 1093–1126. doi:10.1037/pspp0000171

Bruya, B., & Ardelt, M. (2018). Wisdom can be taught: A proof-of-concept study for fostering wisdom in the classroom. *Learning and Instruction*, *58*, 106–114. doi:10.1016/j.learninstruc.2018.05.001

Carlson, E. N. (2016). Meta-accuracy and relationship quality: Weighing the costs and benefits of knowing what people really think about you. *Journal of Personality and Social Psychology*, *111*(2), 250–264. doi:10.1037/pspp0000107

Chakroff, A., Dungan, J., Koster-Hale, J., Brown, A., Saxe, R., & Young, L. (2016). When minds matter for moral judgment: Intent information is neurally encoded for harmful but not impure acts. *Social Cognitive and Affective Neuroscience*, *11*(3), 476–484. doi:10.1093/scan/nsv131

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, *111*(48), 17320–17325. doi:10.1073/pnas.1408988111

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity. *Psychological Bulletin*, *52*(3), 177–193. doi:10.1037/h0044919

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. doi:10.1037/0022-3514.44.1.113

Decety, J. (2011). Dissecting the neural mechanisms mediating mpathy. *Emotion Review*, *3*(1), 92–108. doi:10.1177/1754073910374662

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*(3), 327–339. doi:10.1037/0022-3514.87.3.327

Epley, N., & Waytz, A. (2010). Mind perception. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 498–541). Hoboken, NJ: Wiley. 10.1002/9780470561119.socpsy001014.

Exline, J. J., & Hill, P. C. (2012). Humility: A consistent and robust predictor of generosity. *The Journal of Positive Psychology*, *7*(3), 208–218. doi:10.1080/17439760.2012.671348

Eyal, T., Steffel, M., & Epley, N. (2018). Perspective mistaking: Accurately understanding the mind of another requires getting perspective, not taking perspective. *Journal of Personality and Social Psychology*, *114*(4), 547–571. doi:10.1037/pspa0000115

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, *76*(5), 1267–1316. doi:10.1111/j.1467-6494.2008.00521.x

Gray, K., & Schein, C. (2012). Two minds vs. two philosophies: Mind perception defines morality and dissolves the debate between deontology and utilitarianism. *Review of Philosophy and Psychology*, *3*(3), 405–423. doi:10.1007/s13164-012-0112-5

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. doi:10.1080/1047840X.2012.651387

Grossmann, I., Brienza, J. P., & Bobocel, D. R. (2017). Wise deliberation sustains cooperation. *Nature Human Behaviour*, *1*(3), 0061. doi:10.1038/s41562-017-0061

Gweon, H., Young, L. L., & Saxe, R. (2011). Theory of mind for you, and for me: Behavioral and neural similarities and differences in thinking about beliefs of the self and other. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*(33), 2492–2497.

Kim, M., Park, B., & Young, L. (2020). The psychology of motivated versus rational impression updating. *Trends in Cognitive Sciences*, *24*(2), 101–111. doi:10.1016/j.tics.2019.12.001

Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing. *Human Development*, *26*, 222–232.

Krumrei-Mancuso, E. J. (2017). Intellectual humility and prosocial values: Direct and mediated effects. *The Journal of Positive Psychology*, *12*(1), 13–28. doi:10.1080/17439760.2016.1167938

Krumrei-Mancuso, E. J., & Rouse, S. V. (2016). The development and validation of the comprehensive intellectual humility scale. *Journal of Personality Assessment*, *98*(2), 209–221. doi:10.1080/00223891.2015.1068174

LaBouff, J. P., Rowatt, W. C., Johnson, M. K., Tsang, J.-A., & Willerton, G. M. (2012). Humble persons are more helpful than less humble persons: Evidence from three studies. *The Journal of Positive Psychology*, *7*(1), 16–29. doi:10.1080/17439760.2011.626787

Lee, M., Sul, S., & Kim, H. (2018). Social observation increases deontological judgments in moral dilemmas. *Evolution and Human Behavior*, *39*(6), 611–621. doi:10.1016/j.evolhumbehav.2018.06.004

Lees, J., & Cikara, M. (2020). Inaccurate group meta-perceptions drive negative out-group attributions in competitive contexts. *Nature Human Behaviour*, *4*(3), 279–286. doi:10.1038/s41562-019-0766-4

Lewis, K. L., Hodges, S. D., Laurent, S. M., Srivastava, S., & Biancarosa, G. (2012). Reading between the minds: The use of stereotypes in empathic accuracy. *Psychological Science*, *23*(9), 1040–1046. doi:10.1177/0956797612439719

Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor-observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, *93*(4), 491–514. doi:10.1037/0022-3514.93.4.491

Malloy, T. E., Albright, L., Kenny, D. A., Agatstein, F., & Winquist, L. (1997). Interpersonal perception and metaperception in nonoverlapping social groups. *Journal of Personality and Social Psychology*, *72*(2), 390–398. doi:10.1037/0022-3514.72.2.390

Phillips, J., & Knobe, J. (2018). The psychological representation of modality. *Mind & Language*, *33*(1), 65–94. doi:10.1111/mila.12165

Rakoczy, H., Wandt, R., Thomas, S., Nowak, J., & Kunzmann, U. (2018). Theory of mind and wisdom: The development of different forms of perspective-taking in late adulthood. *British Journal of Psychology*, *109*(1), 6–24. doi:10.1111/bjop.12246

Reeder, G. D., Vonk, R., Ronk, M. J., Ham, J., & Lawrence, M. (2004). Dispositional attribution: Multiple inferences about motive-related traits. *Journal of Personality and Social Psychology*, *86*(4), 530–544. doi:10.1037/0022-3514.86.4.530

Rom, S. C., Weiss, A., & Conway, P. (2017). Judging those who judge: Perceivers infer the roles of affect and cognition underpinning others' moral dilemma responses. *Journal of Experimental Social Psychology*, *69*, 44–58. doi:10.1016/j.jesp.2016.09.007

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*(3), 279–301. doi:10.1016/0022-1031(77)90049-X

Schein, C., & Gray, K. (2016). Moralization and harmification: The dyadic loop explains how the innocuous becomes harmful and wrong. *Psychological Inquiry*, *27*(1), 62–65. doi:10.1080/1047840X.2016.1111121

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–39. doi:10.1177/1088868317698288

Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves. *Social and Personality Psychology Compass*, *4*(8), 605–620. doi:10.1111/j.1751-9004.2010.00280.x

Vazire, S., & Carlson, E. N. (2011). Others sometimes know us better than we know ourselves. *Current Directions in Psychological Science*, *20*(2), 104–108. doi:10.1177/0963721411402478

Waytz, A., Young, L. L., & Ginges, J. (2014). Motive attribution asymmetry for love vs. Hate drives intractable conflict. *Proceedings of the National Academy of Sciences*, *111*(44), 15687–15692. doi:10.1073/pnas.1414146111

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684. doi:10.1111/1467-8624.00304

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, *118*(2), 357–378. doi:10.1037/a0022936

Young, L., & Tsoi, L. (2013). When mental states matter, when they don't, and what that means for morality. *Social and Personality Psychology Compass*, *7*(8), 585–604. doi:10.1111/spc3.12044

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, *19*(4), 399–404. doi:10.1111/j.1467-9280.2008.02099.x