frontiers in PSYCHOLOGY

OPINION ARTICLE
published: 09 April 2013
doi: 10.3389/fpsyg.2013.00171

Undecidability and opacity of metacognition in animals and humans

Kevin B. Clark[1,2]* and Derrick L. Hassert[3]

[1] Research and Development Service, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, USA
[2] Portland, OR, USA
[3] Department of Psychology, Trinity Christian College, Palos Heights, IL, USA
*Correspondence: kbclarkphd@yahoo.com

Edited by:
Mattie Tops, VU University Amsterdam, Netherlands

Metacognition, defined either epistemologically as knowledge about knowledge or operationally as behavior about behavior (Koriat, 2007), presumably enables intelligent agents to self-referentially and, in social contexts (Bahrami et al., 2010, 2012), group-referentially monitor and control emotions, moods, perceptions, memories, reasoning, decisions, and actions. At an epistemological level of description, the nested referent nature of metacognition succumbs to problems originating from recursively enumerable propositional logic; that, as Kurt Gödel (1931) first proved for Bertrand Russell and Alfred North Whitehead's axiomatic *Principia Mathematica*, the meaning of statements created about conditions of a system or set of systems by the respective same system or set of systems can be formally undecidable. Momentarily ignoring peripheral confounds introduced by stochastic and imperfect biological systems, animal and human metacognitive operations and all their possibilities must thus exist in a universe of graded logicomathematical consistency (i.e., All theorems are true syntax-correct propositions of the system.) and completeness (i.e., All true syntax-correct propositions of the system are theorems.) (Kreisel, 1967). Because strong consistency excludes strong completeness, the knotted statements of metacognition may show themselves to be falsehoods, truths, or truths essentially unverifiable in theory as well as in subjective and objective practice (**Figure 1**) (Raattkainen, 2005).

Psychologists well understand the fallibility of formal logic systems and of axiomatic animal and human psychological processes, including, among other phenomena, feature detection, inferential judgments, error diagnosis and correction, concept formation, memory storage and retrieval, and introspection (cf. Nisbett and Ross, 1980; Kahneman et al., 1982; Watanabe and Huber, 2006). We often stipulate—with qualifications—the flawed definition(s) and agent execution of cognition and metacognition (e.g., Shimamura and Metcalfe, 1994; Koriat and Goldsmith, 1996, 1998; Smith, 2009; Terrace and Son, 2009; Frith, 2012; Fleming et al., 2012; Yeung and Summerfield, 2012). For pragmatic reasons, many of us enthusiastic about studying metacognition also avoid the strange, looping causality of self-reference exposed with Gödelian numbering to concentrate on solving basic and/or clinical empirical difficulties that arise when trying to identify this stubbornly opaque hypothetical construct of healthy and pathological minds (e.g., Bach and David, 2006; Koren et al., 2006; Vance, 2006; Carruthers, 2009; Gumley, 2011; Brevers et al., 2013).

The operational opacity of metacognition becomes arguably most apparent when considering: (1) poor subjective accessibility to the cognition and metacognition of animals and humans of limited or no language proficiency (cf. Hampton, 2009; Fleming and Dolan, 2012; Kepecs and Mainen, 2012; Smith et al., 2012), (2) the synergism and antagonism of unreliable explicit and implicit psychological components that mask real metacognitive abilities and capacities from agent and external observer (cf. Hampton, 2009; Fleming et al., 2012), and (3) the occasional independence between metacognition and cognitive skills which may exacerbate the preceding two problems (cf. Koriat and Goldsmith, 1998; Schneider, 1999). A case illustrating these three classes of problems is found for Paulus et al. (2013), who report evidence of implicit metacognition in normal preschool children performing a paired-associates learning and memory task. The authors confront the dilemma of metacognition opacity with a paradigm common to belief, memory, Theory of Mind, and now metacognition research (cf. Nisbett and Ross, 1980; Penn and Povinelli, 2007; Carruthers, 2009; Hampton, 2009; Izard, 2009; Frith, 2012; Skarratt et al., 2012); they prescribe objectively observable, variable primary and secondary behaviors that can be scored for accuracy and/or efficiency and for cross-correlations involving use of secondary behaviors to monitor and control primary behaviors. A typical test scenario has subjects follow associative learning with forced-choice declarative recognition of perceptual/conceptual pairings between two separate images. Subjects subsequently give confidence judgments of memory accuracy through explicit self-reports, such as scalar feelings, and implicit reactions, such as changes in traceable voluntary gaze or involuntary saccades, pupil dilation, pressor effects, and electrical properties of skin. Ratings of accuracy confidence reflect an agent's more-or-less accessible knowledge representation and decisional or post-decisional monitoring and control.

By combining experimental protocols that test for self-reports and behavioral reactions, scientists, including Paulus et al. (2013), hope to objectify a subject's mental states irrespective of his/her language skills as well as dissociate and double dissociate confounding explicit and implicit information processing. For example, Paulus et al. (2013) demonstrate implicit confidence judgments (i.e., gaze direction and duration and pupil area) can be superior in memory accuracy to prompted explicit judgments (i.e., self-report). Increased eye responses coincide with increased (pre)attention demand or load, suggesting humans might be more capable of successfully monitoring judgments at a preattentive or non-conscious level during

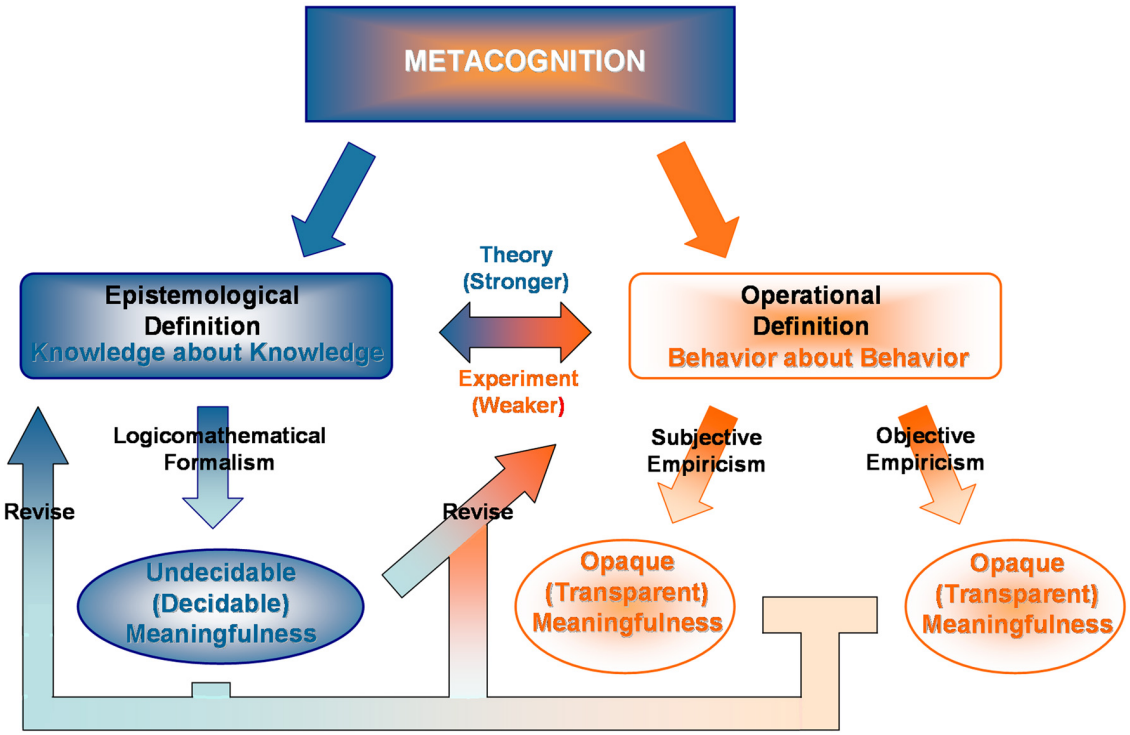

FIGURE 1 | Diagram outlines relationship of metacognition undecidability and opacity with epistemological and operational definitions, where decidability of logicomathematical formalism is a stronger, more global condition than transparency of empiricism. Interestingly, the nested referent structure of epistemological and operational definitions may also lead to the paradox that not only is the meaningfulness of metacognition undecidable, but so too is the meaningfulness of (explicit and implicit) reasoning used to reach that conclusion. Anti-mechanistic, -algorithmic, or -logic views of animal and human intelligence find varying degrees of support from noted thinkers in physics (Penrose, 1994), logic (Kreisel, 1967), philosophy, (Lucas, 1962) and psychology (Kahneman et al., 1982). Though a staunch anti-Intuitionist and pro-mathematical Platonist, Gödel (1951/1995) believed the human mind could overcome with empirical certainty such circumstances as "absolutely unsolvable Diophantine problems" and other undecidable problems suggestive of his theorems.

earlier stages of cognitive development. This sort of approach toward investigating metacognition now seems routine though its widespread appeal dates back roughly 35 years (cf. Flavell, 1979). It builds upon the efforts of Thorndike (1911), Köhler (1927), and additional pioneering psychologists, who inferred mentality in laboratory and wild animals incapable of symbolic communication. Middle to late twentieth century primatologists especially began to embrace behavioral methodology as a tool to compare and contrast the cognitive development and capacity of non-human primates with humans, particularly periverbal neonates, toddlers, and preschoolers (Parker and McKinney, 1999). The value of cross-taxonomic analyses and concomitant realization that oral and written verbal self-reports can be inaccurate, due to lack of awareness and incidental mental events, such as confabulation, illusory perception, cued or irretrievable memory, affect bias, and knowledge lean (e.g., Nisbett and Wilson, 1977; Berry and Broadbent, 1984), triggered this paradigm shift away from once favored subjective human introspection techniques. Some authorities even believe advances in metacognition research are only achievable with improved behavioral models (e.g., Terrace and Son, 2009; Kepecs and Mainen, 2012).

Besides phenomenological and mechanistic insights, studies such as Paulus et al. (2013) on developmental complexity and stages of metacognition across animal and human lifespans figure to clarify the ecological relevance of metacognition, with numerous ramifications for parenting, education, crime, and public health. The weighty private and public consequences of metacognition then necessitate that metacognition be researched and interpreted with exceeding care in regard to cognitive transparency of test paradigms. Debates concerning the effectiveness of standard behavioral tests to reduce metacognition opacity remain spirited at best. Akin to faults of introspection and self-reports (Clark, 2012), behaviors are susceptible to incidental and (pre)attentive disturbances which influence perception, memory, and performance and which may be improperly controlled with experiment conditions (cf. Penn and Povinelli, 2007; Hampton, 2009; Smith, 2009; Smith et al., 2012). Perhaps most alarming, however, are assertions condemning instances of behaviorally measured metacognition in animals and language-challenged humans as nothing more than automatic (i.e., non-reflective), low-level associative response contingencies (Penn and Povinelli, 2007; Hampton, 2009). A conclusion Paulus et al. (2013) also admit could not be negated for some of their observations on the relationship of pupil size, memory retrieval, and metacognition.

This kind of harsh indictment, if correct, despite correlative neurometric findings from non-invasive EEG, PET, fMRI, and SQUID brain recordings (e.g., Fleming and Dolan, 2012), sadly almost reaffirms outcomes of Gödel's Incompleteness Theorems—the nested referential structure of metacognition may prevent metacognition from ever being fully transparent to empiricism and, therefore, makes its emergent meaningfulness perhaps hopelessly undecidable for epistemological and operational definitions (**Figure 1**). However, empiricists tend to believe, as we and Paulus and colleagues do, perfecting the state-of-art for metacognition paradigms gives fair cause for optimism. Gödel (1951/1995) himself considered logicomathematical formalism to exist independently above animal and human mentality. He also denied that his theorems substantiated, with logicomathematical certainty, the concept of humanly unsolvable objective or subjective problems. With irony, he felt this could be known for external or internal (self) observer alone through rational empiricism.

REFERENCES

Bach, L. J., and David, A. S. (2006). Self-awareness after acquired and traumatic brain injury. *Neuropsychol. Rehabil.* 16, 397–414.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., and Frith, C. (2012). What failure in collective decision making tells us about metacognition. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1350–1365.

Bahrami, B., Olsen, K., Latham, P. E., Roepstorff, A., Rees, G., and Frith, C. D. (2010). Optimally interacting minds. *Science* 329, 1081–1085.

Berry, D. C., and Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *Q. J. Exp. Psychol. Sect. A* 36, 209–231.

Brevers, D., Cleeremans, A., Bechara, A., Greisen, M., Kornreich, C., Verbanck, P., et al. (2013). Impaired self-awareness in pathological gamblers. *J. Gaml. Stud.* 29, 119–129.

Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *Behav. Brain Sci.* 32, 121–138.

Clark, K. B. (2012). "A statistical mechanics definition of insight," in *Computational Intelligence*, ed A. G. Floares (Hauppauge, NY: Nova Science Publishers, Inc.), 139–162.

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2213–2223.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *Am. Psychol.* 34, 906–911.

Fleming, S. M., and Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1338–1349.

Fleming, S. M., Dolan, R. J., and Frith, C. D. (2012). Metacognition: computation, biology and function. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1280–1286.

Gödel, K. (1931). Über formal unentscheidbare Säze der Principia Mathematica und verwandter Systeme, I. *Monatsh. Math. Phys.* 38, 173–198.

Gödel, K. (1951/1995). "Some basic theorems on the foundations of mathematics and their implications (Gibbs Lecture)," in *Collected Works, Vol. 3: Unpublished Essays and Lectures*, eds S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, and R. Solovay (Oxford: Oxford University Press), 304–323.

Gumley, A. (2011). Metacognition, affect regulation and symptom expression: a transdiagnostic perspective. *Psychiatry Res.* 190, 72–78.

Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: converging evidence or multiple mechanisms? *Comp. Cogn. Behav. Rev.* 4, 17–28.

Izard, C. E. (2009). Emotion theory: research, highlights, unanswered questions, and emerging issues. *Annu. Rev. Psychol.* 60, 1–25.

Kahneman, D., Slovic, P., and Tversky, A. (eds.). (1982). *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press.

Kepecs, A., and Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1322–1337.

Köhler, W. (1927). *The Mentality of Apes*. New York, NY: Vintage.

Koren, D., Seidman, L. J., Goldsmith, M., and Harvey, P. D. (2006). Real-world cognitive—and metacognitive—dysfunction in schizophrenia: a new appraoch for measuring (and remedying) more "right stuff." *Schizophr. Bull.* 32, 310–326.

Koriat, A. (2007). "Metacognition and consciousness," in *The Cambridge Handbook of Consciousness*, eds P. D. Zelazo, M. Moscovitch, and E. Thompson (Cambridge: Cambridge University Press), 289–325.

Koriat, A., and Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychol. Rev.* 103, 490–517.

Koriat, A., and Goldsmith, M. (1998). "The role of metacognitive processes in the regulation of memory performance," in *Metacognition and Cognitive Neuropsychology: Monitoring and Control Processes,* eds G. Mazzoni and T. O. Nelson (Mahwah, NJ: Lawrence Erlbaum), 97–118.

Kreisel, G. (1967). "Mathematical logic: what has it done for the philosophy of mathematics?" in *Bertrand Russell. Philosopher of the Century,* ed R. Schoeman (London: George Allen and Unwin), 201–272.

Lucas, J. R. (1962). Minds, machines, and Gödel. *Philosophy* 36, 112–137.

Nisbett, R., and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall Inc.

Nisbett, R. E., and Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* 84, 231–259.

Parker, S. T., and McKinney, M. L. (1999). *Origins of Intelligence: The Evolution of Cognitive Development in Monkeys, Apes, and Humans*. Baltimore, MD: The Johns Hopkins University Press.

Paulus, M., Proust, J., and Sodian, B. (2013). Examining implicit metacognition in 3.5-year-old children: an eye-tracking and pupillometric study. *Front. Psychol.* **4**:145. doi: 10.3389/fpsyg.2013.00145

Penn, D. C., and Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling 'theory of mind'. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 731–744.

Penrose, R. (1994). *Shadows of the Mind: A Search for the Missing Science of Consciousness*. New York, NY: Oxford University Press.

Raattkainen, P. (2005). On the philosophical relevance of Gödel's incompleteness theorems. *Rev. Int. Philos.* 4, 513–534.

Schneider, W. (1999). "The development of metamemory in children," in *Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application. Attention and Performance,* eds D. Gopher and A. Koriat (Cambridge, MA: The MIT Press), 487–514.

Shimamura, A. P., and Metcalfe, J. (eds.). (1994). *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press.

Skarratt, R. A., Cole, G. G., and Kuhn, G. (2012). Visual cognition during real social interactions. *Front. Hum. Neurosci.* 6:196. doi: 10.3389/fnhum.2012.00196

Smith, J. D. (2009). The study of animal metacognition. *Trends Cogn. Sci.* 13, 389–396.

Smith, J. D., Couchman, J. J., and Beran, M. J. (2012). The highs and lows of theoretical interpretation in animal metacognition research. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1297–1309.

Terrace, H. S., and Son, L. K. (2009). Comparative metacognition. *Curr. Opin. Neurobiol.* 19, 67–74.

Thorndike, E. L. (1911). *Animal Intelligence: Experimental Studies*. New York, NY: Macmillan.

Vance, D. E. (2006). A review of metacognition in aging with HIV. *Percept. Mot. Skills* 103, 693–696.

Watanabe, S., and Huber, L. (2006). Animal logics: decisions in the absence of human language. *Anim. Cogn.* 9, 235–245.

Yeung, N., and Summerfield, C. (2012). Metacognition in human decision-making: confidence and error monitoring. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1310–1321.

Received: 20 March 2013; accepted: 21 March 2013; published online: 09 April 2013.

Citation: Clark KB and Hassert DL (2013) Undecidability and opacity of metacognition in animals and humans. Front. Psychol. ***4****:171. doi: 10.3389/fpsyg.2013.00171*

This article was submitted to Frontiers in Cognition, a specialty of Frontiers in Psychology.