clearly about the impending rise of robots and help roboticists understand how their creations are likely to be received.

## Acknowledgments

We thank Bertram Malle, Ilan Finkelstein, Michael Clamann, and an anonymous reviewer for comments on a draft of this paper. This work has been supported by the National Science Foundation award SPRF-1714298) to Y.E.B. by the National Science Foundation awards IIS-1149965 and CCF-1533844 to R.A., and a grant from the Charles Koch Foundation to K.G.

[1]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA
[2]Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA
[3]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27514, USA

*Correspondence: ybigman@email.unc.edu (Y.E.E).

https://doi.org/10.1016/j.tics.2019.02.008



## References

1. Awad, E. *et al.* (2018) The moral machine experiment. *Nature*, 563, 59–64
2. Shariff, A.F. *et al.* (2014) Free will and punishment: a mechanistic view of human nature reduces retribution. *Psychol. Sci.* 25, 1563–1570
3. Wegner, D.M. and Gray, K. (2017) *The Mind Club,* Viking
4. Kim, T. and Hinds, P. (2006) Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In *ROMAN 2006: The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 80–85, IEEE
5. van der Woerdt, S. and Haselager, P. (2017) When robots appear to have a mind: the human perception of machine agency and responsibility. *New Ideas Psychol.* http://dx.doi.org/10.1016/j.newideapsych.2017.11.001
6. Bekey, G.A. (2005) *Autonomous Robots: From Biological Inspiration to Implementation and Control,* The MIT Press
7. Weisman, K. *et al.* (2017) Rethinking people's conceptions of mental life. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11374–11379
8. Bigman, Y.E. and Gray, K. (2018) People are averse to machines making moral decisions. *Cognition*, 181, 21–34
9. Kissinger-Knox, A. *et al.* (2018) Does non-moral ignorance exculpate? Situational awareness and attributions of blame and forgiveness. *Acta Anal.* 33, 161–179
10. Monroe, A.E. and Malle, B.F. (2017) Two paths to blame: intentionality directs moral information processing along two distinct tracks. *J. Exp. Psychol. Gen.* 146, 23–33
11. Dudek, G. and Jenkin, M. (2010) *Computational Principles of Mobile Robotics,* Cambridge University Press
12. de Visser, E.J. *et al.* (2016) Almost human: anthropomorphism increases trust resilience in cognitive agents. *J. Exp. Psychol. Appl.* 22, 331–349
13. Waytz, A. *et al.* (2014) The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. *J. Exp. Soc. Psychol.* 52, 113–117
14. Gray, K. *et al.* (2014) The myth of harmless wrongs in moral cognition: automatic dyadic completion from sin to suffering. *J. Exp. Psychol. Gen.* 143, 1600–1615
15. Malle, B.F. *et al.* (2019) AI in the sky: how people morally evaluate human and machine decisions in a lethal strike dilemma. In *Robots and Well-Being* (Aldinhas Ferreira, I., Silva Sequeira, J., Virk, G.S., Kadar, E.E. and Tokhi, O., eds), Springer

## Letter

# Does Explaining Social Behavior Require Multiple Memory Systems?

Pieter Van Dessel [1,*], Bertram Gawronski,[2] and Jan De Houwer[1]

Amodio [1] argues that social cognition research has for many decades relied on imprecise dual-process models that build on questionable assumptions about how people learn and represent information. He presents an alternative framework for explaining social behavior as the product of multiple dissociable memory systems, based on the idea that cognitive neuroscience has revealed evidence for the existence of separate systems underlying distinct forms of learning and memory.

Although we applaud Amodio's attempt to build bridges between social cognition, learning psychology, and neuroscience, we believe that his interactive memory systems model rests on shaky grounds. In our view, the most significant limitation is the idea that behavioral dissociations provide strong evidence for multiple memory systems with functionally distinct learning mechanisms. A major problem with this idea is that behavioral dissociations can arise from processes during the retrieval and use of stored information, which does not require any assumptions about distinct memory systems or distinct forms of learning. For example, in contrast to Amodio's argument that double dissociations between implicit evaluative bias and implicit stereotypical bias in the prediction of different forms of discriminatory behavior provide evidence for distinct memory systems [2], the observed dissociation may simply indicate that people retrieve and use different kinds of information when faced with different kinds of behavioral decisions (e.g., how close to sit next to a stranger vs. whom to choose as a partner for a trivia task). Such differences in the retrieval and use of stored information do not imply that different types of information (e.g., evaluative vs. stereotypical) are stored in distinct memory systems.

The same concern applies to dissociations involving neural structures. For example, in instrumental learning tasks, Parkinson's disease patients with striatal dysfunction have been found to verbally report the correct reward contingencies without making reward-congruent choices, whereas patients with hippocampal lesions show the reversed impairment [3]. Amodio interprets such findings as evidence for independent representations of conceptual and instrumental knowledge arising from distinct forms of learning [1]. However, such dissociations can also arise from differences in retrieval processes drawing upon a single memory system. In line with this concern, it has been argued that dissociations in the behavior of Parkinson's disease and hippocampal lesion patients reflect differences in the expression of a single type of representation in two tasks that require different ways of retrieving these representations [4]. Theoretical ambiguities like these have led to increased skepticism about the idea that cognitive

neuroscience reveals multiple memory systems that are each associated with different neural substrates [5].

Our arguments are also applicable to other dissociations beyond the ones discussed by Amodio. For example, several studies have found that implicit (i.e., spontaneous) evaluations reflect the mere co-occurrence of stimuli regardless of their relation, whereas explicit (i.e., deliberate) evaluations are sensitive to the particular relation of the co-occurring stimuli [6]. Based on extant dual-process theories, such findings have been interpreted as evidence for distinct learning mechanisms underlying implicit and explicit evaluations: automatic formation of associative links between co-occurring events (e.g., associative link between A and B) and controlled generation and truth assessment of mental propositions about the relation between co-occurring events (e.g., A prevents B). However, the observed dissociation may also reflect differences in the retrieval of stored propositional information, given that (i) implicit and explicit evaluations differ in terms of their relative speed and (ii) fast evaluations are more likely affected by incomplete retrieval of stored information (e.g., retrieval of A is related to B rather than A prevents B) [7]. Thus, different from the argument that the observed dissociation provides evidence for functionally distinct learning mechanisms, it can be explained by retrieval-related processes without any assumptions about distinct learning mechanisms or distinct memory systems.

When exploring complexity in the retrieval and use of stored information, social cognition research can draw upon an extensive literature in diverse fields of psychology and neuroscience. For example, a wide range of phenomena such as categorization, task switching, recognition, recall, contingency learning, feature binding, stimulus–response binding, negative priming, and social judgment can be accounted for by episodic memory models that assume a single (episodic) memory system that is operated upon by context-dependent similarity-based retrieval mechanisms [8–10]. Likewise, many complexities of Pavlovian conditioning can be accounted for by assuming a comparator mechanism that compares multiple simple associations at the time of performance [11]. Finally, cognitive neuroscience has seen a surge in the popularity of predictive coding models, which explain a wide range of behavioral findings in terms of highly flexible processes involved in the retrieval and expression of low-level predictions [12]. Social cognition researchers are only beginning to exploit the huge potential that these retrieval-based approaches offer. Following this shift towards explaining behavioral complexity at the level of retrieval might be a more promising way forward for social cognition than a proliferation of learning and memory systems.

## Acknowledgments

Preparation of this article was supported by a Postdoctoral fellowship of the Scientific Research Foundation, Flanders to P.V.D., Methusalem Grant BOF16/MET_V/002 to J.D.H., and by National Science Foundation Grant 1649900 to B.G. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[1]Department of Experimental Clinical and Health Psychology, Ghent University, Belgium
[2]Department of Psychology, University of Texas at Austin, TX, USA

*Correspondence:
Pieter.vanDessel@UGent.be (P. Van Dessel).

https://doi.org/10.1016/j.tics.2019.02.001



## References

1. Amodio, D. (2019) Social Cognition 2.0: an interactive memory systems account. *Trends Cogn. Sci.* 23, 21–33
2. Amodio, D.M. and Devine, P.G. (2006) Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *J. Pers. Soc. Psychol.* 91, 652–661
3. Knowlton, B.J. *et al.* (1996) A neostriatal habit learning system in humans. *Science*, 273, 1399–1402
4. Nosofsky, R.M. and Zaki, S.R. (1998) Dissociations between categorization and recognition in amnesic and normal individuals: an exemplar-based interpretation. *Psychol. Sci.* 9, 247–255
5. Graybiel, A.M. and Grafton, S.T. (2015) The striatum: where skills and habits meet. *Cold Spring Harb. Perspect. Biol.* 7, a021691
6. Corneille, O. and Stahl, C. (2018) Associative attitude learning: a closer look at evidence and how it relates to attitude models. *Pers. Soc. Psychol. Rev.* Published online March 1, 2018. http://dx.doi.org/10.1177/1088868318763261
7. Hu, X. *et al.* (2017) Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on implicit and explicit evaluations. *Pers. Soc. Psychol. Bull.* 43, 17–32
8. Hintzman, D.L. (1986) 'Schema abstraction' in a multiple-trace memory model. *Psychol. Rev.* 93, 411–428
9. Schmidt, J. *et al.* (2016) The parallel episodic processing (PEP) model 2.0: A single computational model of stimulus-response binding, contingency learning, power curves, and mixing. *Cogn. Psychol.* 91, 82–108
10. Smith, E.R. and Zárate, M.A. (1992) Exemplar-based model of social judgment. *Psychol. Rev.* 99, 3–21
11. Stout, S.C. and Miller, R.R. (2007) Sometimes-competing retrieval (SOCR): a formalization of the comparator hypothesis. *Psychol. Rev.* 114, 759–783
12. Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204

Forum

# Obsessive Compulsive Disorder: A Pathology of Self-Confidence?

Julian Kiverstein [1,4,*]
Erik Rietveld [1,2,3,4]
Heleen A. Slagter [4,5,6] and
Damiaan Denys [1,4]

**A striking change OCD patients repeatedly describe following treatment with deep brain stimulation (DBS) of the ventral anterior limb of internal capsule (vALIC) is an immediate increase in self-confidence. We show how the DBS-induced changes in self-confidence reported by our patients can be understood neurocognitively in terms of active inference.**