

situation, and thus none can account for the consistent performances of apes across diverse ToM tests [7]. Specific to false-belief attribution, such evidence even includes the recent finding that apes might understand an experimenter's false beliefs in an interactive helping task [8]. Second, in an experiment based on an earlier proposal by Heyes [5], Karg *et al.* [9] showed that chimpanzees could apply previous self-experience with the occlusive properties of two barriers (i.e., that one was opaque and the other see-through) to determine which path would allow them to steal food from a competitor without being seen, even though at the time of choice the barriers appeared identical and no low-level cues were available to the participants.

Finally, submentalizing could not explain the anticipatory looking of apes in a previous eye-tracking study in which an inanimate control was implemented. Kano and Call [10] tested great apes with movies in which a hand repeatedly reached for and grasped one of two objects. When the locations of the objects were switched and the hand moved centrally toward both, apes looked in anticipation of the hand pursuing a new path to grasp the old goal. However, when watching videos that were identical, except that the hand was replaced with an inanimate mechanical claw, apes looked to both objects equally. They did not anticipate that the claw would pursue the old goal, as they did in the case of the animate agent. Thus, inanimate features of the stimuli could not account for the goal-based action prediction of apes.

In spite of this evidence, we accepted Heyes' [6] challenge and performed an inanimate version of the false-belief task that was highlighted in the author's article. Despite comparable levels of attention, the inanimate stimuli elicited markedly lower anticipatory looking and no significant tendency to look to the correct box (see [11] and Box 1 for a summary). Thus,

evidence from diverse studies – experience–projection, interactive helping, and inanimate controls of implicit goal-understanding and false-belief attribution tasks – converge on the same conclusion: submentalizing is insufficient to explain the social-cognitive abilities of great apes [7–10].

### Acknowledgments

We thank Dr. Cecilia Heyes for discussing the details of our nonsocial submentalizing control experiment reported here.

<sup>1</sup>Kumamoto Sanctuary, Wildlife Research Center, Kyoto University, Kumamoto, Japan

<sup>2</sup>Primate Research Institute, Kyoto University, Inuyama, Japan

<sup>3</sup>Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

<sup>4</sup>School of Psychology and Neuroscience, University of St. Andrews, St. Andrews, UK

<sup>5</sup>Department of Psychology and Neuroscience, Duke University, Durham, NC, USA

\*These authors contributed equally.

\*Correspondence:

fkanou@gmail.com (F. Kano) and

ckrupenye@gmail.com (C. Krupenye).

<http://dx.doi.org/10.1016/j.tics.2017.06.011>

### References

1. Premack, D. and Woodruff, G. (1978) Does the chimpanzee have a theory of mind. *Behav. Brain Sci.* 1, 515–526
2. Southgate, V. *et al.* (2007) Action anticipation through attribution of false belief by 2-year-olds. *Psychol. Sci.* 18, 587–592
3. Krupenye, C. *et al.* (2016) Great apes anticipate that other individuals will act according to false beliefs. *Science* 354, 110–114
4. Heyes, C. (2014) False belief in infancy: a fresh look. *Dev. Sci.* 17, 647–659
5. Heyes, C. (2014) Submentalizing: I am not really reading your mind. *Perspect. Psychol. Sci.* 9, 131–143
6. Heyes, C. (2017) Apes submentalise. *Trends Cogn. Sci.* 21, 1–2
7. Call, J. and Tomasello, M. (2008) Does the chimpanzee have a theory of mind? 30 years later. *Trends Cogn. Sci.* 12, 187–192
8. Buttelmann, D. *et al.* (2017) Great apes distinguish true from false beliefs in an interactive helping task. *PLoS One* 12, e0173793
9. Karg, K. *et al.* (2015) The goggles experiment: can chimpanzees use self-experience to infer what a competitor can see? *Anim. Behav.* 105, 211–221
10. Kano, F. and Call, J. (2014) Great apes generate goal-based action predictions: an eye-tracking study. *Psychol. Sci.* 25, 1691–1698
11. Krupenye, C. *et al.* A test of the submentalizing hypothesis: Apes' performance in a false belief task inanimate control. *Commun. Integr. Biol.* (in press)

## Forum

# Origins of the Belief in Good True Selves

Julian De Freitas,<sup>1,\*</sup>

Mina Cikara,<sup>1</sup>

Igor Grossmann,<sup>2</sup> and

Rebecca Schlegel<sup>3</sup>

**Despite differences in beliefs about the self across cultures and relevant individual differences, recent evidence suggests that people universally believe in a 'true self' that is morally good. We propose that this belief arises from a general tendency: psychological essentialism (PE).**

Beliefs about the concept of a 'self' vary across cultures, perspectives (first versus third), and individuals. Yet mounting evidence suggests that people exhibit a robust, invariant tendency to believe that inside every individual there is a 'good true self' calling them to behave in morally virtuous ways [1]. Where does this belief come from? We propose that it arises from PE: the basic cognitive tendency to assume that all entities have deep, unobservable, inherent properties that comprise their true nature.

The predominant view of PE is that it is the result of several psychological capacities that emerge in early childhood and persist into adulthood. Such capacities include tracking identity continuity and distinguishing appearance from deeper realities [2]. These capacities enable the mind to make better sense of the world. Tracking identity continuity allows recognition of individuals across different contexts, whereas searching for deeper realities allows the making of more accurate predictions by not simply taking surface properties at face value [2].

While most research on PE has examined people's beliefs about categories (e.g.,

animal species, women), there is growing evidence that people apply the same essentialism when reflecting on their own and others' 'selves' [3]. Applying PE to self-beliefs suggests that all individuals have underlying true natures (or self-essences) that constitute their identities. We review three emerging research streams providing initial support to this hypothesis.

First, people equate the good true self with an agent's personal identity. One important property of PE is that removing an entity's seemingly essential characteristics is more consequential for identity than removing seemingly peripheral characteristics. If the seemingly essential characteristics of an entity are removed, it is no longer considered the same entity. Beliefs about morally good true selves exhibit this property: changes in moral characteristics are more disruptive to identity judgments than changes in other personality characteristics or morally neutral or immoral characteristics [4,5]. For example, if a person is described to have undergone a transplantation, people are more likely to say that the person is no longer the same person if the transplantation leads to a change in their moral characteristics compared with their emotions, autobiographical memories, or lower-level cognitive traits [4]. Similarly, people are more likely to attribute change from immoral to moral behavior (e.g., deadbeat dad to involved dad) to the emergence of the person's true self compared with analogous changes from moral to immoral behavior or morally neutral changes in preferences (e.g., a change in one's favorite sport) [5]. This research indicates that people believe that moral characteristics make up the true self; thus, the good true self appears to be viewed as most essential to a person's identity.

Second, true-self beliefs exhibit hallmark features of PE. People believe that self-essences, like category essences, are stable and naturally occurring (rather

than artificial). Specifically, people rate personality traits that they deem central to a person's identity as more 'innate' and stable over time relative to other traits [6]. People also believe that there is a boundary between the self-essence and other aspects of the self, since they spontaneously describe the true self as a physical entity 'inside' or 'beneath the surface' (of the extrinsic self) that can 'grow', 'expand', or be 'expressed' [7]. Relatedly, self-essences are believed to have non-obvious properties; people believe that private thoughts are more indicative of the true self than public actions [8]. Finally, self-essences are believed to be diagnostic of what is true about an individual, such as whether they are happy or blameworthy or value something [9]. For example, people believe that an immorally acting person who expresses signs of happiness is not really happy; she must feel unhappy 'deep down' given that she is betraying her presumed 'good nature'. However, when the same person is explicitly described as having an immoral true self, she is rated as happier when acting immorally because she is believed to be expressing her 'bad nature' [9].

Finally, belief in a good true self may be universal. Belief in a good true self: (i) has been observed in independent and several interdependent cultures that vary in affluence; (ii) applies equally to both ingroup and outgroup members; and (iii) is evident even among misanthropes [5]. In all of these studies, people either explicitly or implicitly endorsed the notion that morally good characteristics comprise the self-essence whereas immoral characteristics do not. This resilience across boundary conditions again suggests that PE, which operates similarly across cultures and various individual differences [10], drives belief in a good true self.

In each of these potential boundary conditions, respondents could have been more concerned with immoral features

or could have failed to distinguish between the true self and other parts of the self, but they did not. For instance, it is well known that cultures vary in the extent to which the self is viewed as independent versus interconnected with others. In independent cultures (e.g., the USA), focusing on private, inner attributes might lead to self-enhancement, whereas reflection on the same attributes may lead to self-criticism in interdependent cultures [11]. Independent cultures also view such inner attributes as more critical to the self-concept, whereas interdependent cultures emphasize less-essentialist attributes concerning relations with significant others [12]. Despite such established differences in self-judgments, belief in a good true self was documented in both the USA and three markedly different interdependent cultural groups – Colombia, Russia, and Singapore. For all three groups, people were more likely to say that the true self caused a change towards moral behavior than neutral or immoral behavior, although the groups varied socioeconomically and made substantially more interdependent self-characterizations than in the USA [5].

Although preliminary, empirical evidence so far seems to suggest that belief in a good true self is culturally universal. Notably, even if this belief is universal it is still possible that the way cultures elaborate on this belief varies. For example, socialization and culture-bound factors such as religion and parenting might provide the content of the beliefs (i.e., what particular behaviors are considered morally good) while the mind provides the form in which the content is represented. Until we find evidence that people in these cultures appeal to observable individual characteristics or other factors over and above deeper qualities when explaining their own and others' behavior, or treat immoral or non-moral characteristics as more essential than moral characteristics, the current self-essence hypothesis stands.

Overall, emerging evidence suggests that belief in a good true self is rooted in psychological essentialism. Although any one of the features reviewed above is tentative, they tend to cohere, making the essentialism account parsimonious. The boundary conditions we have tested thus far failed to disconfirm the priority and robustness of belief in a good true self. Future work should continue exploring boundary conditions, including tests of clinical samples, and more remote cultures. We caution that each piece of evidence will need to be reconciled with the separate aspects of the belief, including its structural features, content (the essence concerns the moral domain), and valence (the essence is morally good). For now, we conclude that

psychological essentialism may explain why, despite robust differences in beliefs about the self in other domains, there is a consistent propensity to believe that each and every one of us possesses a good true self.

<sup>1</sup>Harvard University, Cambridge, MA, USA

<sup>2</sup>University of Waterloo, Waterloo, Ontario, Canada

<sup>3</sup>Texas A&M University, College Station, TX, USA

\*Correspondence: [defreitas@g.harvard.edu](mailto:defreitas@g.harvard.edu) (J. De Freitas).  
<http://dx.doi.org/10.1016/j.tics.2017.05.009>

#### References

1. Strohminger, N. et al. The true self: a psychological concept distinct from the self. *Perspect. Psychol. Sci.* (in press)
2. Gelman, S.A. (2003) *The Essential Child: Origins of Essentialism in Everyday Thought*, Oxford University Press
3. De Freitas, J. et al. (2016) Normative judgments and individual essence. *Cogn. Sci.* 41, 1551–6709
4. Strohminger, N. and Nichols, S. (2014) The essential moral self. *Cognition* 31, 159–171
5. De Freitas, J. et al. Consistent belief in a good true self in misanthropes and three interdependent cultures. *Cogn. Sci.* (in press)
6. Haslam, N. et al. (2004) Essentialist beliefs about personality and their implications. *Pers. Soc. Psychol. Bull.* 30, 1661–1673
7. Moser, K.S. (2007) Metaphors as symbolic environment of the self: how self-knowledge is expressed verbally. *Curr. Res. Soc. Psychol.* 12, 151–178
8. Johnson, J.T. et al. (2004) Inferences about the authentic self: when do actions say more than mental states? *J. Pers. Soc. Psychol.* 87, 615–630
9. Newman, G.E. et al. (2015) Beliefs about the true self explain asymmetries based on moral judgment. *Cogn. Sci.* 39, 96–125
10. Atran, S. (1990) *Cognitive Foundations of Natural History*, Cambridge University Press
11. Heine, S.J. and Hamamura, T. (2007) In search of East Asian self-enhancement. *Pers. Soc. Psychol. Rev.* 11, 4–27
12. Kung, F.Y. et al. (2016) Culture, fixed-world beliefs, relationships, and perceptions of identity change. *Soc. Psychol. Pers. Sci.* 7, 631–639