



Reports

The duality of virtue: Deconstructing the moral hypocrite

Piercarlo Valdesolo*, David DeSteno

Department of Psychology, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

ARTICLE INFO

Article history:

Received 2 July 2007

Revised 2 January 2008

Available online 30 March 2008

Keywords:

Morality

Decision making

Cognitive control

ABSTRACT

The present study examined the processes giving rise to moral hypocrisy, a phenomenon in which individuals judge their own transgressions to be less morally objectionable than the same transgressions enacted by others. Two alternative models of the source of hypocrisy were compared to determine whether hypocrisy results from automatic or volitional biases. Findings demonstrated not only that participants viewed their own transgressions as significantly more “fair” than the same transgressions enacted by others, but also that this bias was eliminated under conditions of cognitive constraint. These findings support the view that hypocrisy stems from volitionally-guided justifications, and thereby suggest that at a more basic level, humans possess a negative response to violations of fairness norms whether enacted by themselves or others.

© 2008 Elsevier Inc. All rights reserved.

Introduction

It's vile. It's more sad than anything else, to see someone with such potential throw it all down the drain because of a sexual addiction.

- Former Congressman Mark Foley on Bill Clinton, 1998, before resigning amidst allegations of sexual misconduct in 2006.

Moral hypocrisy refers to a fundamental bias in moral judgment in which individuals evaluate a moral transgression enacted by themselves to be less objectionable than an identical transgression enacted by others (Valdesolo & DeSteno, 2007). Of high import for intergroup relations, this asymmetric leniency has been shown to extend to others as a function of their relation to the self: a transgression enacted by a member of an ingroup is perceived to be of equal acceptability to the same transgression enacted by the self, but to be more acceptable than the identical behavior enacted by an outgroup member or non-affiliated other (Valdesolo & DeSteno, 2007). Although on first blush this finding may seem somewhat unsurprising for groups characterized by long-standing conflict (e.g., Israeli vs. Palestinian factions), its value lies in its demonstration among emergent groups. That is, moral hypocrisy readily arises even when using minimal groups, thereby attesting to the deep-seated nature of the bias (Valdesolo & DeSteno, 2007).

Given both its apparent elemental status and practical import, moral hypocrisy stands as a phenomenon quite worthy of further investigation. At present, the existence of moral hypocrisy is clear,

but the mechanisms that underlie it remain clouded. Accordingly, the present experiment focuses on examining the process(es) by which moral hypocrisy emerges.

Uncovering the hypocritical mind

To elicit hypocrisy, we developed a paradigm in which individuals faced a dilemma representing a conflict between self-interest and the interest of another (Valdesolo & DeSteno, 2007; cf. Batson, Thompson, Seufferling, Whitney, & Strongman, 1999). In this paradigm (to be described in more detail below), some participants were required to divide a resource (i.e., expended time and energy) between themselves and another, and could do so either fairly (i.e., through a random allocation procedure) or unfairly (i.e., through personal selection of the preferred option). They were later asked to evaluate the morality, or fairness, of their actions. Other participants viewed a separate individual, who was a confederate, acting in the unfair manner toward another (i.e., selecting the better option for herself) and subsequently evaluated the morality of this act. We defined hypocrisy as the discrepancy between the fairness judgments for this same transgression when committed by the self or by the other.

By modeling hypocrisy as discrepant moral judgments, we might expect that its underlying mechanisms would operate in a fashion similar to that of any other moral evaluation. Recent research in the psychology of morality has begun to converge on a dual-process model of moral judgment (Cushman, Young, & Hauser, 2006; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Haidt, 2001; Pizarro & Bloom, 2003; Valdesolo & DeSteno, 2006). According to this view, an intuitive process is theorized to work in tandem with more domain-general, consciously-guided processes to mediate decision making. Processes at both levels are sensitive,

* Corresponding author. Present address: Department of Psychology, Amherst College, Amherst, MA 01002, USA.

E-mail addresses: piercarlo.valdesolo@gmail.com (P. Valdesolo), d.desteno@gmail.com (D. DeSteno).

to differing degrees, to morally-relevant events or principles (e.g., cause no direct harm, utility, self-protection), with the eventual decision output representing some confluence of the processes. We believe that moral hypocrisy can be understood within this framework.

Conceptualizing hypocrisy as a dual-process model, however, leads to competing predictions regarding precisely how these two classes of processes interact to produce the phenomenon. More specifically, hypocrisy could be driven by a discrepancy in automatic intuitions in response to one's own versus another's transgressions. That is, individuals might display an automatic positivity bias for their own transgressions relative to others', with higher order processes simply functioning to create post hoc justifications for "gut-level" decisions (cf. Haidt, 2001). Alternatively, hypocrisy might be driven by differential activation of higher order cognitive processes geared toward justification and rationalization of one's own transgressions. That is, although individuals might have negative automatic reactions to both their own and others' transgressions, they may engage in more consciously motivated reasoning when judging their own transgressions in order to maintain a positive self-view.

Distinguishing between these two competing explanations has important practical implications for developing strategies geared toward curbing this disturbingly familiar phenomenon. Indeed, deciding whether intuitions should be fostered or overcome hinges upon whether or not people have automatic aversions to their own as well as others' violations of fairness norms.

Two alternative models

As noted, there is reason to believe hypocrisy could emerge in two ways based on a dual-process model of moral judgment. Mounting evidence suggests that humans may have evolved an intuitive aversion to violations of equity, with similar aversions evidenced by certain primate species (Brosnan & de Waal, 2003; Hauser, 2006). It has also been hypothesized that humans have evolved specific social emotions designed to foster cooperation and trust with others (Bartlett & DeSteno, 2006; Frank, 1988; Haidt & Graham, 2007), suggesting an important role for emotional responses designed to inhibit self-serving behavior and, thereby, to avoid negative social consequences. Accordingly, violations of fairness stand as a strong candidate to engender a spontaneous and immediate negative reaction regardless of the enactor, suggesting that hypocrisy might emerge from more deliberative processes.

Similarly, several lines of research suggest that higher order processes might be employed to rationalize and justify a self-enacted transgression (Bandura, 1990; Bandura, 1996; Haidt, 2001; Trivers, 1985). In this case, the intuitive system would favor a more "moral" judgment in accord with a basic fairness norm (i.e., showing self-interest is not appropriate), but conscious control systems might work to generate a more "immoral" judgment (i.e., showing self-interest is permissible) that nevertheless may serve to protect one's self-image. However, when judging another's transgression, higher order processes should not temper the intuitive response as the motive for self-image preservation is not relevant.

Alternatively, recent findings demonstrate that disruption of brain regions involved in cognitive control can decrease aversion to inequity within the context of economic games (Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006), suggesting that automatic reactions might be geared toward engendering self-serving, as opposed to fair, behavior. Indeed, this finding aligns with much research suggesting that humans possess an automatic positivity bias with respect to evaluations involving the self. For instance, tests of implicit self-esteem consistently reveal a seemingly ubiquitous generalized positive evaluation of self (Greenwald & Farnham, 2000; Yamaguchi et al., 2007). In a similar vein, much work

has suggested that exaggerated perceptions of mastery and unrealistic optimism are characteristic of normal human thought (Taylor & Brown, 1988). When taken in combination with recent research demonstrating that both motivational states (Balcetis & Dunning, 2006) and chronic views regarding one's abilities (Ehrlinger & Dunning, 2003) are capable of influencing low-level automatic processes, these findings suggest that chronic views of oneself as a moral individual, as well as motives to appear as such, might lead to positively biased spontaneous evaluations of one's own transgressions relative to those of others.

If it is the case that the intuitive system does not generate an immediate aversion, or at least a lesser one, to an individual's own transgressions, then hypocrisy might simply arise as a result of discrepant, spontaneous evaluative responses. According to this view, the intuitive system would favor a more "moral" judgment in accord with a basic fairness norm when contemplating other's transgressions, but favor a more "immoral" judgment in accord with an automatic positivity bias when contemplating one's own. Put simply, individuals might not be as sensitive to transgressions that bring one immediate benefits. If true, these intuitions would work in concert with higher order processes which would serve to provide post hoc explanations for the behavior.

The present experiment

The present experiment seeks to disentangle these competing explanations. If hypocrisy derives from competition between a negative affective response to any violation of fairness coupled with conscious efforts aimed at justifying the behavior when enacted by oneself, then hypocrisy should disappear when efforts aimed at conscious control are constrained (cf. Greene et al., 2004; Valdesolo & DeSteno, 2006). However, if hypocrisy arises because of discrepant automatic intuitions generated in response to one's own versus another's transgressions, then constraining conscious control should have no effect on judgments of the morality of one's own transgressions.

To examine this question, we used a factorial design, crossing judgments of self- and other transgressions with a manipulation of cognitive constraint: a 2 (Enactor: Self vs. Other) \times 2 (Constraint: Control, Cognitive Load). In the control conditions, we expected to replicate the usual hypocrisy effect identified by Valdesolo and DeSteno (2007): participants who acted immorally (i.e., violated the fairness norm) should judge their own fairness transgression to be less objectionable than the same transgression enacted by another. Of import, however, we also expected that reduced ability for controlled processing would alter the relative causal force of processes contributing to judgment, directly addressing the nature of the dual mechanisms underlying hypocrisy. If manipulation of cognitive constraint has no influence on judgments of participants' own transgressions, it would suggest that the model is one wherein hypocrisy arises from biased automatic intuitions. However, if increased cognitive constraint results in more "moral" judgments of participants' own transgressions (i.e., one's own actions are judged to be more unfair) and thereby attenuates hypocrisy, these findings would suggest that hypocrisy arises from discrepant volitional efforts aimed at justifying transgressions when enacted by the self-relative to others. We expected that the manipulations of cognitive constraint would not influence participants' judgments of the confederate's transgressions, as motivated reasoning processes should not be engaged when judging violations committed by neutral others. Consequently, conditions involving the judgments of others will function not only as a baseline for computation of the hypocrisy measure, but also to show that any effects of the manipulations do not represent global influences on moral decision making (e.g., increased cognitive constraint decreases the perceived fairness of any actions, whether enacted by the self or another).

Method

Participants

Ninety-one individuals (58 females, 33 males) participated and were randomly assigned to one of four experimental conditions.

Procedural overview

As the load condition procedures constitute minor variants of the control condition procedures, the procedures for the two primary control conditions will be described in detail, with descriptions of the other conditions limited to noting the small differences in design. In all conditions, participants judged the fairness of an identical action, which served as the primary dependent variable. As noted, we employed a 2×2 design crossing judgments of self- and other transgressions with cognitive load. Presentation of all materials and data collection were accomplished using MediaLab software.

Control conditions

Condition 1: Judging one's own transgression

On entering the lab, a participant was seated at an individual workstation, given a brief introduction to the experiment, and told to begin the computerized tasks. The instructions explained that the experimenters were examining performance on two different types of tasks, and that any participant would only complete one of the tasks. The first task (i.e., the green task) consisted of a brief survey combined with a short photo hunt that would take 10 min to complete. The second task (i.e., the red task) consisted of a series of math and logic problems combined with a longer and somewhat tedious mental rotation task that would take 45 min to complete. Following the task descriptions, participants were informed by the experimenter that the research team was also evaluating a new participant assignment protocol meant to reduce experimenter bias. Therefore, certain participants would be randomly selected to make condition assignments for themselves and others.

Participants then read the following instructions:

In order for the experimenters to remain blind to condition assignments, you must assign either yourself or the next participant to the green condition and the other of you to the red condition. Some people feel that giving both individuals an equal chance is the fairest way to assign the tasks.

If you would like to use a randomizer to assign conditions, please move to the computer behind you and follow the instructions. The decision is entirely up to you. You can assign yourself and the other participant however you choose. The other participant does not and will not know that you are assigning conditions.

The randomizer was a computer program designed to assign the participant to the "red" condition following a few demonstration trials conducted by the experimenter in which it alternated between conditions to guard against participant suspicion. The experimenter then left the room to allow the participant to make his or her choice.

After assigning conditions to themselves and another, participants responded to questions regarding the assignment procedure which were presented as a way to collect opinions on the new protocol. Embedded in a small set of distractors was the target question: "How fairly did you act?" Participants responded to this question on a 7-point scale ranging from "extremely unfairly" to

"extremely fairly". The session was then terminated and participants debriefed.

Condition 2: Judging another's transgressions

In this condition, participants' primary task involved evaluating the actions of another individual who completed a procedure identical to the one completed by the participant in Condition 1. Here, participants were informed that their role was to act as an impartial observer to provide feedback to experimenters regarding use of the new assignment protocol by other participants. These other participants were in fact confederates. To accomplish this goal, participants were informed that they would be seated in the room with an individual taking part in an experiment and therefore able to observe his actions and responses to the experimental protocol through the use of a yoked computer. That is, participants would be able to see on their screen what the other participant was reading and selecting in real time. Participants received the following instructions on their screen:

Your computer is connected to the adjacent computer. Another participant will be completing an experiment on that computer and you will be asked to follow along and observe on your screen everything that he reads and does. Note that the other participant will be unaware that this is happening. After approximately 5 min of observing, you will be asked to rate the new assignment protocol in terms of clarity and design as well as answer some questions concerning the performance of the participant.

Participants were asked if they understood their task, and if so to click the mouse to connect the two computers. From this point on, they were presumably observing the other participant's screen and were asked not to touch their computer until it disconnected and automatically moved them along to the evaluations.

After the computers had "connected," the participant waited in her seat while the experimenter brought in the second participant (i.e., the confederate). The confederate was told that all instructions would be on the computer and to begin the experiment by clicking the mouse. The confederate then simultaneously clicked his mouse as well as a second mouse surreptitiously connected to the back of the participant's computer. The mouse clicks set off a timed presentation which created the illusion that the participant was observing, on her own monitor, the confederate go through the instructions and assign himself the "green" condition and a future participant the "red" condition without using the randomizer. After observing the confederate's choice, the participant's computer "disconnected" and brought her to an evaluation section where, embedded in a set of distractors, she answered the following target question: "How fairly did the participant act?" using the same scale as in Condition 1.

Cognitive constraint conditions

Condition 3: Judging one's own transgression

Condition 3 was a replication of Condition 1 with the exception that participants made fairness judgments under cognitive load. The load manipulation came directly after participants assigned tasks to themselves and the other, thereby affecting only moral judgment and not behavior. Cognitive load was manipulated using a digit-string memory task (cf. Gilbert & Hixon, 1991). Participants were told that the experimenters were interested in how people make judgments when they are distracted. To simulate distraction, they would be asked to remember a string of digits at the same time that they were responding to a series of questions. Participants were told that a string of seven digits would appear on the

screen before each question. They would then have to answer the question within 10 s, immediately after which they would have to recall the digit string that had preceded the question. Participants were also told that it was extremely important to provide the most accurate answers possible for questions comprising the assignment evaluation measure. The primary dependent variable consisted of the fairness question presented and scaled as in Condition 1 and embedded in the series of distractor questions completed under load.

Condition 4: Judgments of another's transgressions

Condition 4 mirrored Condition 2 with the exception that participants made judgments under cognitive load, using the same load manipulation as in Condition 3.

Results

Participants in conditions involving judgments of their own transgressions were removed from analysis if they did not commit a transgression. That is, only those participants who assigned themselves the “green” (i.e., preferable) condition and who did not use the randomizer were included in the analysis. As in previous research, those who immediately acted either altruistically or in accord with the fairness norm were a substantial minority (cf. Valdesolo & DeSteno, 2007); this group consisted of 7 (8%) participants spread almost equally across the two relevant conditions (i.e., Conditions 1 and 3).

Moving to the full factorial design, an ANOVA confirmed the predicted interaction between the Enactor and Cognitive Constraint factors, $F(1,87) = 6.20$, $p = .02$, $d = .75$ (see Fig. 1). As expected, moral hypocrisy emerged in the control conditions; the same fairness transgression was judged to be substantially more moral when enacted by the self than when enacted by another, $t(49) = 3.39$, $p = .001$, $d = .95$. However, constraints on effortful correction (i.e., cognitive load) resulted in the disappearance of

the hypocrisy effect; participants experiencing load judged their own transgressions to be as unfair as the same behavior when enacted by another, $t(38) = .12$, $p = .91$. Indeed, a planned contrast revealed that judgments of one's own actions in the control condition (i.e., Condition 1) significantly exceeded judgments in any of the other three conditions, which showed no reliable differences among themselves, $F(1,87) = 5.12$, $p = .003$.¹

Discussion

The present study provides strong evidence that moral hypocrisy is governed by a dual-process model of moral judgment wherein a prepotent negative reaction to the thought of a fairness transgression operates in tandem with higher order processes to mediate decision making. Hypocrisy readily emerged under normal processing conditions, but disappeared under conditions of cognitive constraint. Inhibiting control prevented a tamping down or override of the intuitive aversive response to the transgression. Of import, these findings rule out the possibility that hypocrisy derives from differences in automatic affective reactions towards one's own and others' transgressions. Rather, when contemplating one's own transgression, motives of rationalization and justification temper the initial negative response and lead to more lenient judgments. Motivated reasoning processes are not engaged when judging others' violations, rendering the prepotent negative response more causally powerful and leading to harsher judgments.

These findings are also noteworthy for demonstrating that controlled processing need not always function to “correct” more basic, intuitive responses (cf. Greene et al., 2004), but rather can be subject to less admirable motives such as the protection of self-image. Indeed, they show that the interplay between intuitive and volitional moral reasoning is sensitive not only to abstract moral principles but also to more selfish motivations, as evidenced by the overwhelming majority of participants who acted unfairly when assigning tasks.²

Despite this disconcerting result, the unearthing of a prepotent negative response to one's own transgressions, and conversely the absence of an automatic positivity bias, reveals an adventitious relationship between moral judgment and hypocrisy. The detection of a low-level sensitivity to fairness transgressions, even at the cost of one's own potential short-term gain, adds to the growing body of evidence dispelling theories which describe morality as a tenuous and fragile “veneer” laid over a core of selfish impulses (c.f. de Waal, 2006). Instead, it seems likely that humans have evolved strong intuitions which, though selected to promote long term self-interest via reciprocal altruism, can represent moment-to-moment instances of pure selfless concern. Yet, our hypocritical behavior belies this intuition. In light of such findings, future work should aim to further define the conditions which temper hypocrisy

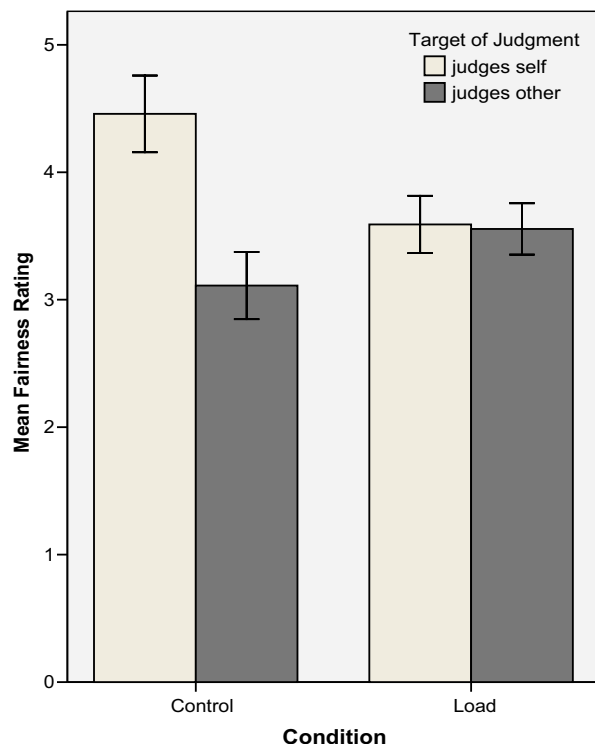


Fig. 1. Mean fairness judgments as a function of condition. Error bars indicate ± 1 SE.

¹ *t*-Tests involving all intergroup comparisons were in the predicted direction. Participants in Condition 1 judged their transgressions to be more fair than participants in any other condition (all *t*'s > 2.28, all *p*'s < .03).

² Lest one wonder whether the ubiquity of fairness transgressions signaled a belief that the utilized fairness transgression was not viewed as immoral, we collected data from a separate sample of 42 participants who were presented with the dilemma facing our participants (i.e., whether to use the randomizer or assign oneself the easy task). We asked these participants to indicate whether they believed a choice based on self-interest was moral or immoral. All 42 identified it as immoral. This finding replicates work by Batson in which 78 of 80 participants in a study utilizing a methodology similar to our own (Batson, Kobryniewicz, & Dinnerstein, 1997) reported that the most “morally right” way to assign tasks was either by assigning the other participant the preferable task or by using a random method of task assignment. Consequently, although such transgressions are not viewed as strongly immoral (i.e., falling slightly below the midpoint on the fairness scales used in this study), it is clear that such self-interested actions clearly fall into the immoral realm.

and ultimately suggest ways in which humans can better translate moral feelings into moral actions.

References

- Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*(4), 612–625.
- Bandura, A. (1990). Selective activation and disengagement of moral control. *Journal of Social Issues*, *46*(1), 27–46.
- Bandura, A. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, *71*(2), 364–374.
- Bartlett, M. Y., & DeSteno, D. A. (2006). Gratitude and prosocial behavior: Helping when it costs you. *Psychological Science*, *17*(4), 319–325.
- Batson, C. D., Kobryniewicz, D., & Dinnerstein, J. L. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology*, *72*(6), 1335–1348.
- Batson, C. D., Thompson, E. R., Seufferling, G., Whitney, H., & Strongman, J. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, *77*, 525–537.
- Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, *425*, 297–299.
- Cushman, F. A., Young, L., & Hauser, M. D. (2006). The role of reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089.
- De Waal, F. B. M. (2006). *Primates and philosophers: How morality evolved*. Princeton: Princeton University Press.
- Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*(1), 5–17.
- Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions*. New York: Norton.
- Gilbert, D. T., & Hixon, G. J. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, *44*, 389–400.
- Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*(6), 1022–1038.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*, 98–116.
- Hauser, M. D. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York: Harper Collins.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832.
- Pizarro, D. A., & Bloom, P. (2003). The intelligence of moral intuitions: Comment on Haidt (2001). *Psychological Review*, *110*, 197–198.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193–210.
- Trivers, R. L. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.
- Valdesolo, P., & DeSteno, D. A. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*, 476–477.
- Valdesolo, P., & DeSteno, D. A. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science*, *18*, 689–690.
- Yamaguchi, S., Greenwald, A. G., Banaji, M. R., Murakami, F., Chen, D., Shiomura, K., et al. (2007). Apparent universality of positive implicit self-esteem. *Psychological Science*, *18*, 498–500.