The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

The SPOT Effect: People Spontaneously Prefer their Own Theories

Aiden P. Gregg, Nikhila Mahadevan & Constantine Sedikides

To cite this article: Aiden P. Gregg, Nikhila Mahadevan & Constantine Sedikides (2016): The SPOT Effect: People Spontaneously Prefer their Own Theories, The Quarterly Journal of Experimental Psychology, DOI: 10.1080/17470218.2015.1099162

To link to this article: http://dx.doi.org/10.1080/17470218.2015.1099162

Accepted author version posted online: 02 Feb 2016.

Publisher: Taylor & Francis & The Experimental Psychology Society

Journal: *The Quarterly Journal of Experimental Psychology*

DOI: 10.1080/17470218.2015.1099162

Running head: SPOT EFFECT

The SPOT Effect:
People Spontaneously Prefer their Own Theories

Aiden P. Gregg

Nikhila Mahadevan

Constantine Sedikides

University of Southampton

Corresponding author: Aiden P. Gregg, Centre for Research on Self and Identity, Psychology, University of Southampton, Southampton, SO17 1BJ, England, UK. Email: aiden@soton.ac.uk. This research was supported in part by Grant IH 113 from the Thrive Center for Human Development, funded by the John Templeton Foundation.

Abstract

People often exhibit confirmation bias: they process information bearing on the truth of their theories in a way that facilitates their continuing to regard those theories as true. Here, we tested whether confirmation bias would emerge even under the most minimal of conditions. Specifically, we tested whether drawing a nominal link between the self and a theory would suffice to bias people towards regarding that theory as true. If, all else equal, people regard the self as good (i.e., engage in self-enhancement), and good theories are true (in accord with their intended function), then people should regard their own theories as true; otherwise put, they should manifest a *Spontaneous Preference for their Own Theories* (i.e., a *SPOT effect*). In three experiments, participants were introduced to a theory about which of two imaginary alien species preyed upon the other. Participants then considered in turn several items of evidence bearing on the theory, and each time evaluated the likelihood that the theory was true versus false. As hypothesized, participants regarded the theory as more likely to be true when it was arbitrarily ascribed to *them* as opposed to an "Alex" (Experiment 1) or to no one (Experiment 2). We also found that the SPOT effect failed to converge with four different indices of self-enhancement (Experiment 3), suggesting it may be distinctive in character.

In general, people see themselves favorably. This phenomenon, called *self-enhancement*, takes many forms (for reviews, see: Gregg & Sedikides, 2014; Sedikides & Gregg, 2008). Amongst other things, people exhibit a "self-enhancing triad" (Taylor & Brown, 1988): more than strict rationality would warrant, they entertain favorable self-views (Schmitt & Allik, 2005), consider themselves to be in control (Presson & Benassi, 1996), and regard their future prospects as bright (Helweg-Larsen & Shepperd, 2001). But perhaps the most blatant sign of self-enhancement is social comparative in nature: the *better-than-average effect* (*BTAE*; Alicke & Govorun, 2005).

Self > Others I: The Better-Than-Average Effect

People generally judge themselves to stand better than their peers across a variety of everyday desirable dimensions. For example, 90% of US road users put themselves in the top 50% in terms of their driving ability (Svenson, 1981); 50% of US academics put themselves in the top 10% as regards their teaching ability (Cross, 1977); and 25% of US high-school students put themselves in the top 1% as regards their social ability (College Board, 1976–1977). Even incarcerated criminals consider themselves nicer than the "average community member" and no less law-abiding (Sedikides, Meek, Alicke, & Taylor, 2014). Such strikingly self-serving judgments suggest that they are shaped by a potent motive to self-enhance (Alicke & Sedikides, 2009, 2011).

Admittedly, the BTAE is not entirely motivationally based. As critics have contended (Hamamura, Heine, & Takemoto, 2007; Moore, 2007), several cognitive factors also underlie it. These include a general preference for individuals over collectives (Klar & Giladi, 1997), an overweighting of information about the self relative to others (Pahl & Eiser, 2005), and the fact that the dimensions judged are often commonplace and controllable (Kruger, 1999). Nonetheless, the BTAE effect persists even when controlling for key confounds (Alicke, Vredenburg, Hiatt, & Govorun, 2001) and is moderated in ways that are difficult to explain only cognitively (Guenther & Alicke, 2010; for a review, see Sedikides & Alicke, 2012). For example, above-average effects covary with the personal importance of the dimension judged, whether it is measured or manipulated (Brown, 2010), and across both Western and Eastern cultures (Sedikides, Gaertner, & Cai, 2015; Sedikides, Gaertner, & Toguchi, 2003).

In sum, the balance of evidence implies that people see themselves as better than average at least partly because they want to.

Importantly, the BTAE is also contagious: not only are self-evaluations susceptible to comparative inflation, but so too are evaluations of entities *connected* to the self. For example, reflecting the fact that people identify with significant others (Aron et al., 2005), parents regard their own children as possessing more positive and fewer negative attributes than children in general, and they do so in proportion to how favorably they regard themselves (Wenger & Fowers, 2008; see also Buunk, 2001). Similarly, reflecting the fact that people identify with larger collectives (Searle, 1995), people show evaluative and behavioral preferences for everyday groups to which they belong over groups to which they do not (Hewstone, Rubin & Willis, 2002), and again in proportion to how favorably they regard themselves (Otten, 2004).

Self > Others II: Ingroups, Names, Birthdates, & Possessions

Even though people are often unaware of the fact that they exhibit a BTAE (Pronin & Kugler, 2007), it nonetheless requires an explicit comparison of self with others, or alternatively, separate explicit evaluations of self and others, which the researcher then compares. Yet the relative inflation of entities connected to the self also emerges in more subtle ways (Greenwald & Banaji, 2005).

For example, arbitrarily assigning people to different groups is sufficient to make them discriminate positively in favor of fellow group members when distributing rewards (Tajfel, Billig, Bundy, & Flament, 1971). This *minimal group discrimination effect* is consistent, amongst other things, with the self acting as an evaluative base, from which greater positivity is assigned to an ingroup (Gramzow & Gaertner, 2005; Otten, 2005). Recent evidence suggests, moreover, that a preference for the ingroup emerges, among those who identify with it, even when the ingroup and outgroup differ in name only, and even when those preferences are indirectly assessed (Roth & Steffens, 2014).

Inanimate entities also succumb to such comparative inflation. For example, people manifest a spontaneous preference for letters in their own name over letters in other people's names, even controlling for valence-relevant confounds such as letter frequency (Koole &

Pelham, 2003; Nuttin, 1987). This *name letter effect* emerges for the digits that enumerate one's birthday (Kitayama & Karasawa, 1997). Moreover, such lexical and numerical egoism has consequences. Participants told that a disreputable historical character (Rasputin) shared the same birthday as they did subsequently denigrated him less (Finch & Cialdini, 1989); and participants expecting to interact with someone assigned a code featuring their own birthday digits anticipated a better interaction (Jones, Pelham, Carvallo, & Mirenberg, 2004). Even pairing one's own (as opposed to another person's) name with some arbitrary symbol is sufficient to make people like that symbol more (Feys, 1991).

A further sign of such transferred egoism is the *endowment effect*: people demand substantially more money to relinquish an existing possession than they offer to acquire that same possession in the first place (Carmon, & Ariely, 2000; Kahneman, Knetsch, & Thaler, 1990). Although loss aversion (Tversky & Kahneman, 1991) or transaction costs (Mandel, 2002) may partly explain the effect, recent research implicates personal ownership as critical. In particular, the effect disappears if buyers own duplicate possessions or if proxy sellers do not own any (Morewedge, Shu, Gilbert, & Wilson, 2009), making the endowment effect a type of *mere ownership effect* (Beggan, 1992). Indeed, people generally describe their own possessions as having more positive, and fewer negative, attributes than other people's possessions, including when they, or another people, come to own a possession by random assignment (Nesselroade, Beggan, & Allison, 1999). Moreover, bolstering or threatening self-regard, thereby setting in motion self-enhancement dynamics, respectively weakens or reinforces the endowment effect (Chatterjee, Irmak, & Rose, 2013), and narcissists believe that their own prized possessions are more positively distinctive than independent raters do (Lee, Gregg, & Park, 2013). Taken together, these findings suggest that the self imparts a motivationally-rooted Midas-like touch to personal possessions that increases their subjective value. Yet even the bare act of choosing one object over another—thereby establishing a differential link to self—is sufficient to induce a preference for that object at a basic associative level (Gawronski, Bodenhausen, & Becker, 2007).

In sum, several lines of research converge on the following conclusion: people are prone to regard themselves and what they own as superior to others and what they own.

Self > Others III: Theories?

Given that comparative self-inflation occurs across several domains, one might also expect it to apply to *theories* that people hold. How precisely? Whatever other merits a theory may have—such as intellectual ingenuity or aesthetic appeal—it will be of greater value if it is *correct* rather than incorrect, given that the primary function of any theory, as opposed to fictional tale, is to represent faithfully some aspect of the world. All else equal, then, a good theory will be true, and a bad theory false. If so, it further follows that, to the extent that one evaluates the self and entities connected to it positively, and to the extent that a theory qualifies as an entity connected to the self because one has originated or endorsed it, one should be inclined to regard that theory *as more likely to be true than false*. That is, people's reasoning processes should be biased by the following heuristic, expressible in the form of a syllogism: If what I am and what is mine is good, and if good theories are true theories, then my theories should be true too.

The operation of such a heuristic is consistent with the existence of the well-established phenomenon of *confirmation bias* (Nickerson, 1998), whereby people process information bearing on the truth of their theories in a way that facilitates their continuing to regard those theories as true. As with the BTAE, there is little doubt that many of the antecedents of confirmation bias are cognitive in character (Klayman & Ha, 1987; Wason, 1968); but, once again, substantial evidence exists for a motivational underpinning (Dunning, 2014; Kunda, 1990). For example, prior ideological commitments, such as political prejudices, bias the reception of arguments (Taber & Lodge, 2006); so too do serious personal consequences, such as the results of medical tests convey (Ditto, Munro, Apanovich, Scepansky, & Lockhart, 2003). Moreover, the degree of biased processing in favor of one's own theories is also surprisingly independent of a key cognitive factor: IQ (Stanovich, West, & Toplak, 2013).

However, as the subtle forms of self-inflation reviewed above indicate, even nominal links between the self with various entities—groups, names, numbers, and goods—suffice to enhance the evaluation of those entities. Might it also be the case, therefore, that even a nominal link between the self and some theory would suffice to enhance evaluations of that

theory, such that it would be judged more likely to be true? In particular—in the absence of any ideological commitments or personal consequences—could drawing an arbitrary connection between self and some theory produce a confirmation bias, as evidence for or against that theory accumulated? Otherwise put, would people show a spontaneous preference for their own theories—a *SPOT effect*—under the most minimal of conditions? If so, then mere ownership of a theory, of the most superficial and nominal sort, would be shown to be enough to occasion a type of confirmation bias.

A thematic precedent is worth remarking upon here. *Mnemic neglect* is a processing bias in spontaneous recall whose operation implicates a motive to self-enhance. Specifically, people are inclined, as a form of strategic self-defence, to forget negative feedback selectively —but only when it pertains to traits they consider important an opposed to trivial, and only when those traits are ascribed to oneself as opposed to someone else (Sedikides & Green, 2009). In the standard case, participants are led to believe that they are receiving real feedback, based on responses to a questionnaire (their own or another person's). However, mnemic neglect emerges even when participants receive purely *hypothetical* feedback (Green, Sedikides, & Gregg, 2008; Sedikides & Green, 2000)—that is, even when they are asked to imagine that they are receiving negative feedback about important (vs. unimportant) traits directed towards them (vs. another person). Thus, even the slightest supposition of self-involvement suffices to induce motivated forgetting. We wondered whether a similarly subtle effect obtained when it came to spontaneous preferences for endorsing theories.

Overview

To test for the SPOT effect, we described a fantasy scenario in which, on a faraway planet, two species of alien existed. The theory at issue was that one of the species was a predator, whereas the other one was its prey. The theory was initially stated, accompanied by a piece of evidence bearing on it. Participants were then asked to rate the likelihood of the theory being true or false. Thereafter, additional pieces of evidence were each presented in turn, and, after each one, participants completed the same rating task. An arbitrary link to self was (or was not) created by altering the wording of the theory's presentation, such that it was

introduced and then repeatedly described as being (or as not being) the participant's own theory.

In Experiment 1, we tested whether the theory was regarded as more likely to be true when arbitrarily ascribed to oneself than to another concrete person. In Experiment 2, we tested whether the theory was regarded as more likely to be true when arbitrarily ascribed to oneself than when either arbitrarily ascribed to another concrete person or to no one at all. In Experiment 3, we explored whether the magnitude of the SPOT effect was related to various other prima facie indices of self-enhancement.

Experiment 1

Method

Platform and procedure. The experiment, neutrally entitled "Theory and Evidence," was created using the online software *iSurvey*™ (University of Southampton, 2015), and distributed online via the crowdsourcing service *CrowdFlower*™ to adult residents of English-speaking countries. Participants indicated their consent, completed the experiment, read a debriefing statement, and entered a voucher number to obtain a pre-arranged nominal payment ($0.10), given the brevity of their participation.

Participants and screening. We initially included a total of 406 cases, identified by *iSurvey* as survey completions. Of these, we conservatively excluded 18.5% on one or more of the following grounds: (a) having duplicate IP addresses (3.4%); (b) completing the experiment too quickly (3.7%), defined as less than half the median completion time; and (c) failing to complete all the truth likelihood estimates (12.6%). All reported at least good English, so none were excluded on that ground.

The final screened sample comprised 331 participants, of whom 38.7% were male. Its members varied in age (M_{AGE} = 34.4; SD_{AGE} = 12.3) and education (56.5% with at least a university degree, the remainder with at least a high school or equivalent diploma), and came largely from North America (USA: 86.6%; Canada: 6.1%; UK: 3.3%; all others: 4.0%). Screened and unscreened samples yielded equivalent results, so we report only the former.

Measures and manipulation.

Demographics. Prior to completing the experiment, participants reported their gender (*male*, *female*), age (in years), the country where they currently lived (from a list of 196), their fluency in English (*perfect*, *very good*, *okay*, *bad*, *very bad*), and their level of education (*higher degree*, *college degree*, *intermediate degree*, *secondary level*, *primary level*).

The Wugworld scenario. Following standard admonitions to answer honestly, participate seriously, and read questions carefully—combined with assurances that there were no right or wrong answers, and that anonymity would be preserved—participants were introduced to the experimental task. Up front, they were instructed that their primary goal was to "estimate how likely it is that a particular theory [was] either true or false." They were then asked to imagine that some party—let us symbolize them as "P"—was "researching life on a planet in a distant solar system" called "Wugworld." This planet was described as inhabited by many alien creatures, some of which were predators, and others prey. The terms "predators" and "prey" were, moreover, mutually defined to ensure comprehension. Specifically, participants were told that "predators […] hunt, kill, and eat prey" whereas "prey […] are hunted, killed, and eaten by predators." Participants were further told that two of the creatures on Wugworld were called "Niffites" and "Luupites," and that P was "interested in researching them."

P, it transpired, had already discovered a fact about Niffites and Luupites. Labelled "1," and offset in bold and italics for emphasis, it read that "Niffites are at least twice as large as Luppites." Participants were then asked to"[s]uppose [P has] a theory, based on the fact above." Also offset in bold below, the theory read as follows: "Niffites hunt, kill, and eat Luupites." Next, participants were asked to "[…] indicate, on the basis of the fact provided, how likely it is that [P's] theory is either true or false, using the scale below." The scale consisted of a moveable slider, where participants could, by clicking and dragging with their computer mouse, shift a bar horizontally towards either end of a bipolar scale, marked only by the contrasting endpoints "Certain to be FALSE" and "Certain to be TRUE." Specifically, participants were told "Move the slider towards the right if you think your theory is more likely to be TRUE;" "Move the slider towards the left if you think your theory is more likely to be FALSE;" and "Place the slider near the middle if you think your theory is about equally

likely to be TRUE or FALSE." The slider initially appeared in the middle of scale. Moving the slider returned a position-sensitive value ranging between 0 (*Certain to be FALSE*) and 100 (*Certain to be TRUE*).

Once participants had provided their truth likelihood estimate, they advanced to the subsequent screen. Thereafter, six additional screens, identical in format and featuring the same slider, appeared. Each screen had P discovering "a new fact" about the alien creatures. The subsequent three facts, labelled "2," "3," and "4," were designed to provide somewhat circumstantial support for the theory. They read as follows: "Niffites have powerful teeth and dangerous spikes on their heads;" "The number of Niffites in a region rises and falls with the number of Luupites in a region;" and "Niffites and Luupites often have aggressive interactions." The final three facts, labelled "5," "6," and "7," were designed to cast more definitive doubt on the theory. They read as follows: "Luupites have sharper teeth than Niffites do;" "Luupites have been observed eating the dead bodies of Niffites;" and Luupites have been observed attacking Niffites." On each of the six screens, participants were instructed to provide each new likelihood estimate "on the basis of the facts provided so far."

The experimental manipulation was straightforward: *P was either the participant themselves or another person*. When P was the participant, the word "you"—addressing them directly—appeared in the Wugworld text; when P was the other person, the word "Alex" appeared instead. Accompanying verbs and possessive forms were also modified. In addition, to rule out a methodological confound, we experimentally counterbalanced which set of alien creatures, Niffites or Luupites, was said by the theory to be the putative predator or its prey, by switching around the names. These were the only differences between the experimental and control condition.

Note that several features of the Wugworld scenario were engineered to help test our experimental hypothesis. First, the theory that participants considered—which pertained to life on an imaginary planet far away, and posited a relation between two previously unknown aliens—could hardly be one towards which they would have any pre-existing partisan bias. This permitted us to test cleanly for the comparative inflation of theories connected to the self, in the absence of any complications such as prior ideological commitment. Second, the

names for the aliens, Niffites and Luupites, were carefully pretested in other research to ensure their semantic nullity and evaluative neutrality (Gregg, Seibt, & Banaji, 2006). This served to discourage the formation of connotations suggestive of predators or prey. Finally, the theory was ascribed in a control condition to a concrete person, going by the popular and androgynous name of "Alex" (cf. Sedikides & Green, 2009). This had two benefits: (a) it arguably served to provide a more stringent test of our hypothesis, given that self-enhancement effects are typically smaller when contrasted against a concrete other (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995); and (b) it ruled out the possibility that ascription per se, rather than ascription to the self, was the key factor, compared to a control condition where a theory was left unascribed.

Results and Discussion

The estimated likelihood that the theory was true served as the dependent variable in a 2 ×2 ×7 mixed Analysis of Variance (ANOVA) design, which featured Ascription Target (self vs. other) and Alien Name (Niffites = predators vs. Luupites = predators) as the two between-group factors, and Fact Number (facts 1 through 7) as the single within-group factor.

As anticipated, a significant and substantial main effect of Fact Number emerged, $F(6, 1962) = 386.49$, $p < .0001$, $\eta^2 = .74$. Inspection of the pattern of means suggests that estimated truth likelihood initially rose, then plateaued, next declined sharply, and finally slightly levelled off. Consistent with this observation, a significant linear trend emerged, $F(1, 327) = 752.45$, $p < .0001$, $\eta^2 = .70$, alongside both a quadratic trend, $F(1, 327) = 278.87$, $p < .0001$, $\eta^2 = .46$, and a cubic trend, $F(1, 327) = 70.54$, $p < .0001$, $\eta^2 = .18$. Accordingly, participants were generally responsive to the earlier facts tending to confirm, or at least maintain, the target theory, and the later facts tending to refute it.

Did the ascription of the theory to the participant or to Alex influence estimates of the theory's likely truth in the predicted direction? It did (Figure 1). A significant main effect, averaging across all within-subject levels, emerged, $F(1, 327) = 8.33$, $p < .005$, $\eta^2 = .03$, $d = .32$. In addition, a significant interaction emerged between Ascription Target and Fact Number, $F(6, 1962) = 2.25$, $p = .04$, $\eta^2 = .01$, implying that not all facts reflected the impact of the manipulation equivalently. In particular, estimates did not significantly differ for facts

1 and 2 individually, but did for all subsequent facts, suggesting that impact of the Theory Ascription registered only as evidence accumulated. However, Alien Name neither exerted a main effect, $F(1, 327) = .52$, $p = .47$, nor interacted with Ascription Target, $F(1, 327) = 0.01$, $p = .93$. Moreover, no other interactive effects emerged.

Although we lacked any theoretical grounds for suspecting that age or gender would moderate the SPOT effect, we nonetheless undertook relevant analyses to ensure generality. For convenience, we dichotomized participants into young and old based on a median split (excluding two participants not declaring their age; *Mdn* = 31). We then conducted a 2 x 2 ANOVA, which featured Ascription Target (self vs. other) and Age Group (young vs. old) as between-subjects factors. As expected, an effect for Ascription Target emerged, $F(1, 325) = 8.13$, $p < .005$, $\eta^2 = .02$, but it did not interact with Age Group, $F(1, 325) = .002$, $p = .96$. In a parallel analysis, gender (excluding one participant who did not declare it) also did not moderate the SPOT effect, $F(1, 326) = 1.48$, $p = .22$.

Thus, the results of Experiment 1 supported our hypothesis. Participants regarded the theory as more likely to be true when it was arbitrarily ascribed to them as opposed to someone else, regardless of their age or gender. That is, they exhibited a SPOT effect, spontaneously preferring their own theories. This can be understood as a confirmation bias driven by a mere ownership effect, in which the positive valence of the self is imparted to a theory by sheer association.

Experiment 2

Experiment 2 elaborated upon Experiment 1 in three key ways. First, it sought to replicate the SPOT effect under somewhat varied conditions. In particular, three features of the experiment changed. First, alternative names for the alien creatures were adopted in order to ensure that the originals were not necessary to the effect. Second, the participants, instead of completing an online survey in private as individuals, now filled out survey booklets in public as members of a group (though still anonymously). These varied conditions provided a modest test of the methodological generalizability of the effect. Third, and most important, Experiment 2 added a condition in which the theory was ascribed to *no one*. This enabled us to test whether ascribing a theory to the self was the crucial factor inclining people to regard

that theory as more likely to be true. The results of Experiment 1, for example, could have been due either to a spontaneous preference *for* one's own theories, or to a spontaneous prejudice *against* another person's theories. Experiment 2 permitted these possibilities to be disambiguated. Specifically, we predicted that the self-ascription condition would differ from both the other-ascription and no-ascription conditions, but that these conditions would not differ from one another, thereby showing that the SPOT effect was specific to the self.

Method

Platform and procedure. Participants took part as members of audiences whom the experimenter addressed. Each audience consisted of between 30 and 60 members. Members were, in different groups, either undergraduate students currently enrolled at the University of Southampton or prospective undergraduate students on visit days accompanied by their parents. In both cases, running the experiment served a joint pedagogical and data-collection purpose, acting as a springboard to further instruction and discussion, revolving around the debriefing. Participation was voluntary and unpaid. Consent was negotiated verbally, with audience members being assured that, if they preferred, they could simply observe proceedings. The experiment was essentially run by having participants fill out, with the experimenter's verbal guidance, 8-page booklets that reflected the various experimental conditions. Prior to administration, these booklets had been arranged into random piles in order to ensure non-systematic distribution to audience members.

Participants and screening. A total of 400 individuals participated. Of these, we conservatively excluded 5.5% on one of two grounds: (a) failing to complete all responses constituting the dependent variable (0.8%) or (b) showing zero variance on the dependent variable (5.0%). (All were native speakers of English.) The screened sample of 378 participants consisted of two subsamples: 121 current undergraduates and 275 prospective undergraduates and their parents. Both were predominantly female—87.6% and 77.3% respectively—but the latter was a decade older on average (M_{AGE} = 20.7 versus M_{AGE} = 29.9), as well as more chronologically diverse (SD_{AGE} = 3.9 versus SD_{AGE} = 15.7) owing to its bimodal distribution, with unsurprising peaks around the ages of both 18 and 50. Screened and unscreened samples yielded equivalent results, so we report only the former.

Measures and manipulation.

***Demographics*.** Prior to completing the experiment, participants reported their gender (*female, male*) and age (in years).

Modifications to the scenario. Given that the experiment was administered by booklet, additional admonitions were provided: to read each page in strict sequence and not to confer with neighbours. The content of the Wugworld scenario was, however, almost identical to that in Experiment 1, apart from a few minor textual alterations. Individual facts were presented, and individual likelihood estimates recorded, no longer on successive screens, but on successive pages of the booklet. The form taken by the dependent measure, however, was non-trivially adjusted, to assist both participants making estimates and coder recording their responses. In particular, the slider that participants dragged left or right in Experiment 1 was replaced by a horizontal line that they instead marked with their pens. This line was subdivided into segments defined by tick marks. Half the tick marks denoted percentages ranging from 0% to 100% in increments of 10%; the other half denoted unlabelled percentages lying halfway in between. As before, the leftward and rightward endpoints were labeled "Certain to be false" (0%) and "Certain to be true" (100%) respectively; but additionally the midpoint (50%) was labelled "equally likely to be true or false", and was flanked by intermediate labels "More likely to be false" (25%) and "More likely to be true" (75%). Instructions were adapted accordingly. Coders, with the aid of rulers, later attempted to derive a number from 0 to 100 from where participants placed their marks; however, most respondents put marks at, rather than between, ticks, so most likelihood estimates recorded were divisible by either 5 or 10.

In terms of experimentally intended changes, the two alien creatures were blandly renamed "Dassites" and "Fommites," again with the intention of ridding them of semantic content. Counterbalancing proceeded as before. The ascription of the theory to self or other also proceeded as before via the insertion of either "you have/your" or "Alex has/Alex's" into the text. However, in the condition where no theory was ascribed to no one, a slightly rephrasing was necessary. We substituted "Suppose *there is* a theory…" and "…that *the* theory is true or false…" for the nominative and possessive respectively.

Results and Discussion

The estimated likelihood that the theory was true again served as the dependent variable, this time in a 3 ×2 ×2 ×7 mixed ANOVA design, which featured Ascription Target (self vs. other vs. no-one), Alien Name (Dassites = predators vs. Fommites = predators), and Audience Type (enrolled vs. visiting) as the three between-subjects factors, and Fact Number (facts 1 through 7) as the single within-subjects factor.

As before, a significant and substantial main effect of Fact Number emerged, $F(6, 2196) = 276.94$, $p < .0001$, $\eta^2 = .43$. Estimates of the theory's likelihood of being true followed a similar but not identical trajectory as before: it initially rose, then seesawed up and down, declined sharply, and finally slightly levelled off. As before, a significant linear trend emerged, $F(1, 366) = 502.45$, $p < .0001$, $\eta^2 = .58$, alongside a quadratic trend, $F(1, 366) = 166.11$, $p < .0001$, $\eta^2 = .31$, and a cubic trend, $F(1, 366) = 32.36$, $p < .0001$, $\eta^2 = .08$. Once again, the facts presented tended to initially confirm, then subsequently refute, the theory in the minds of participants, more or less as intended. In addition, Fact Number unexpectedly interacted with Audience Type, $F(6, 2196) = 4.40$, $p < .0001$, $\eta^2 = .01$, such that enrolled audiences gave higher truth likelihood estimates initially, but lower ones subsequently, than visiting audiences—probably a spurious result.

Did the ascription of the theory to the participant, to Alex, or to no one, influence estimates of the theory's likely truth? It did (Figure 2). A significant main effect, averaging across all within-subjects levels, emerged, $F(1, 366) = 3.59$, $p = .03$, $\eta^2 = .02$. Unlike in Experiment 1, no significant interaction emerged between Ascription Target and Fact Number, $F(6, 2196) = 1.20$, $p = .28$. But as before, Alien Name neither exerted a main effect, $F(1, 366) = 0.93$, $p = .76$, nor interacted with Ascription Target, $F(1, 366) = 0.93$, $p = .40$. In addition, no other main or interactive effect reached significance, except for a barely interpretable four-way interaction, $F(12, 2196) = 2.28$, $p < .01$.

Collapsing across all other factors, we proceeded to test more specifically, using between-subjects planned contrasts, whether the manipulation of theory ascription exerted its effects. Averaging across all within-subjects levels, we first contrasted the self condition against the other and no-one conditions jointly. As predicted, estimates of the likely truth of the theory were significantly greater in the self condition, $t(375) = 2.85$, $p < .0005$, $d = .34$. Also as predicted, the other and no-one conditions did not differ significantly from one another, $t(375) = 0.16$, $p = .88$. Moreover, estimates in the self condition individually exceeded those in the other condition, $t(375) = 2.37$, $p < .02$, $d = .29$, and in the no-one condition, $t(375) = 2.57$, $p < .02$, $d = .31$.

As a supplementary analysis, we again tested the generality of the SPOT effect across different ages and gender. Inspection of the age distribution revealed a marked bimodality, with one range spanning 15 to 28 years ($n = 270$) and another 39 to 65 years ($n = 100$). Accordingly, we dichotomized participants into young and old based on either range membership, excluding age-undeclared participants ($n = 8$). In addition, for the sake of simplicity and statistical power, we collapsed the other and no-one conditions into a single condition, to create a dichotomy between non-self and self. We then conducted a 2 x 2 ANOVA, which featured Ascription Target (self vs. non-self) and Age Group (young vs. old) as between-subject factors. As expected, an effect for Ascription Target emerged, $F(1, 366) = 10.19$, $p < .005$, $\eta^2 = .03$. However, it did not interact with Age Group, $F(1, 366) = 1.52$, $p = .22$. In a parallel analysis, gender (excluding six participants who did not declare it) also did not moderate the SPOT effect, $F(1, 368) = 2.38$, $p = .12$.

Thus, the results of Experiment 2 again supported our hypothesis. Participants regarded the theory as more likely to be true when it was arbitrarily ascribed to them as opposed to either someone else or to no one at all, regardless of age or gender.

Experiment 3

Experiment 3 attempted to replicate the SPOT effect for a third time, but had two additional goals.

The first goal was to check whether being called Alex oneself, or having a friend called Alex, might have significantly weakened the results, given that either possibility could

have in principle disposed participants to evaluate an Alex-ascribed theory more favorably. Nonetheless, we considered this unlikely, because: (a) although Alex is a relatively popular name (making the top 100 names circa 2015: http://www.babycenter.com/baby-names-alex-462505.htm), the proportion of participants in the control condition named Alex was still likely to be very small; and (b) the impact of a theory ascribed to a friend is liable to be less than the impact of a theory ascribed to oneself, given the converging evidence for the relative primacy of the individual self (Sedikides, Gaertner, Luke, O'Mara, & Gebauer, 2013) and for people having more positive associations towards themselves than towards even close others (Gebauer, Göritz, Hofmann, & Sedikides, 2012).

The second goal was to test whether and to what extent the SPOT effect varied as a function of four prima facie indices of self-enhancement. Two took the form of traits: (i) *narcissism*, a grandiose form of self-regard, and (ii) *deceptive self-enhancement* (with respect to being right), a tendency to overestimate one's abilities. Two more took the form of manifest biases: (i) *overclaiming* (i.e., claiming to recognize bogus words, as opposed to real ones) and (ii) *biased argument evaluation* (i.e., judgments of argument quality that are idiosyncratic as opposed to consensual). By using standard measures of well-established traits commonly deemed indicative of excessive self-enhancement, as well as more direct measures of self-enhancement in operation, we hoped to cover all our bases, so to speak. On the one hand, a priori reasoning, and empirical evidence, suggest that egoistic evaluations of self-relevant entities should be related to other indices of self-evaluation (Gramzow & Gaertner, 2005; Wenger & Fowers, 2008). On the other hand, observed correlations are sometimes unexpectedly small. For example, various studies have found that the correlation between name letter preferences and narcissism ranges from r = -.08 to r = 0.17 (Bosson et al., 2008), possibly due to psychometric or other defects in the measurement of the former (Hoorens, 2014). Consequently, our expectations here were tentative.

Method

Platform and procedure. In Experiment 3, the Wugworld scenario and additional indices of self-enhancement were administered as part of a larger online study lasting for approximately an hour. The Wugworld scenario appeared at the beginning of this study to

avoid possible priming effects. The larger study was created, presented, and distributed in the same way as Experiment 1, although for appropriately higher pay ($3.00).

Participants and screening. We screened Experiment 3 participants as part of the larger online study rather than as participants in Experiment 3 alone. We recorded a total of 1918 survey attempts by *iSurvey*. Of these, we classified1575 (82.1%) as genuine (more than half the survey complete). From this subset, we conservatively excluded another 54.2% on one or more of the following possible or actual grounds: (a) reporting poor or very poor English (0.0%); (b) having duplicate IP addresses (35.7%); completing the experiment too quickly (3.7%), defined as less than half the median completion time; (c) failing to complete at least 95% of the survey (4.2%); (d) showing zero variance on any survey measure where some variance was expected (26.4%). The high rate of IP duplication was due to unexpected smallness of the crowdsourced population combined with having been obliged to run the survey in several independent stages, thereby permitting repeated participation that we discovered belatedly. The high zero variance policy, though it probably excluded many good participants, also maximized data quality.

The final screened sample consisted of 722 participants, of whom 38.9% were male. Its members varied reasonably in age (M_{AGE} = 36.0; SD_{AGE} = 11.7) and education (53.8% with at least a college degree, and all but two of remainder with at least a high school or equivalent diploma), and came largely from North America (USA: 47.5%; Canada: 18.4%) and other English-speaking countries (UK: 26.6%; Ireland: 1.8%; Australia, 1.8%; New Zealand, 0.8%; all others: 3.1%). Screened and unscreened samples yielded equivalent results, so we report only the former.

Measures and manipulation.

Demographics. Participants reported the same demographics as in Experiment 1.

The Wugworld scenario. The presentation and content were identical to Experiment 1 except that group names were not counterbalanced across predators and prey. This was a simplification justified by the absence of any prior effects and the desirability of single version of the SPOT for correlational investigations. In addition, following presentation of the scenario, participants were asked two questions, to which they answered “yes” or “no” by

selecting the appropriate response from a drop down menu: "Is your own name Alex?" and "Do you have a friend called Alex?"

Narcissism. We assessed narcissism using the 16-item abbreviation of the Narcissistic Personality Inventory (Ames, Rose, & Anderson, 2006), whose psychometric properties approximate the original scale. Our version (α = .80), like other recent precedents (cf. Lee et al., 2013), featured bipolar items that were not dichotomous, but continuous along seven-point scales, for greater reliability. Participants responded by shifting a moveable slider, originally at the midpoint, left or right. Sample item: "When people compliment me, I get embarrassed (R) [vs.] Everybody keeps telling me how good I am."

Self-deceptive enhancement (judgment). We took seven items from the self-deceptive enhancement subscale of the Balanced Inventory of Desirable Responding, Version 7 (Paulhus, 1998) which specifically assessed egoistical overestimations of the accuracy of one's judgment. Participants responded by clicking on radio buttons that defined 7-point scales, ranging from "completely false" to "completely true." Sample items: "My first impressions of people usually turn out to be right," "I am a completely rational person." To maximize internal consistency, we excluded two items with low item-total correlations before computing the total (α = .62)

Overclaiming. We pioneered a version of the overclaiming technique (Paulhus & Harms, 2004) suited for internet administration, given that online participants can easily "Google" the veracity of alleged claims. Following practice trials, participants were shown 50 words for three seconds each. Of these, one set of 25 were real but relatively obscure (e.g., *borborymous*, *wimple*), whereas the other set of 25 were bogus but comparably real-looking (e.g., *challometry*, *hylocenium*). Participants were instructed to click a button only if they recognized a word as real, and otherwise to permit the survey to advance to the next screen automatically. Overclaiming was operationalized as the number of fake words participants claimed to recognize divided by the number real words they claimed to recognize.

Biased Argument Evaluation. We administered an updated version of the Argument Evaluation Test (Gregg, Mahadevan, & Sedikides, 2015; Stanovich & West, 1997), designed to be briefer, more standardized, and less US-centric version of the original, featuring entirely

new content. It also incorporated a different criterion of normative rationality, relying on consensus rather than expertise. Participants began by indicating, on 7-point scales whose options were represented by radio buttons, their level of agreement or disagreement with 24 assertions about a variety of contentious political issues (e.g., "The institution of marriage is meant for one man and one woman," "The government should ban the selling of high-calorie drinks in large containers"). Having provided their own opinions, they then read, for each assertion, a set of three additional statements: an argument for it, a counterargument against that argument, and a rebuttal to that counterargument. In each case, participants were then instructed to judge, again on 7-point scales, "how weak or strong […] the rebuttal […] is, while ignoring your own opinion." On completion of the data collection, and after screening the data, mean ratings of the quality of the 24 rebuttals were computed. These consensual judgments operationalized the objective quality of those rebuttals.

Thereafter, an idiosyncratic simultaneous regression was computed for each participant. In it, participants' own opinion about each of the 24 assertions served as one independent variable; consensual judgments of quality of each of the 24 rebuttals served as the other independent variable; and the participants' judgments of the quality of each of the 24 rebuttals served as the dependent variable. In line with Stanovich and West (1997), the beta weight expressing the independent prediction of each participant's judgment by the consensual judgment was taken as the primary index of the ability to evaluate arguments independently of prior beliefs. We inverted this index to capture biased argument evaluation.

Results and Discussion

The estimated likelihood that the theory was true served as the dependent variable in a 2 ×7 mixed ANOVA design, which featured Ascription Target (self vs. other) as the between-subjects factors, and Fact Number (facts 1 through 7) as the within-subjects factor.

Once again, a significant and substantial main effect of Fact Number emerged, $F(6, 4296) = 612.98$, $p < .0001$, $\eta^2 = .64$. Truth likelihood estimates followed the same pattern as in Experiment 1, exhibiting significant trends that were linear, $F(1, 716) = 1067.22$, $p < .0001$, $\eta^2 = .60$, quadratic, $F(1, 716) = 636.72$, $p < .0001$, $\eta^2 = .47$, and cubic, $F(1, 716) = 134.99$, $p < .0001$, $\eta^2 = .16$. As in Experiment 1, a significant main effect of Ascription

Target emerged (Figure 3) in the predicted direction, $F(1, 716) = 9.37$, $p < .005$, $\eta^2 = .01$, $d = .23$, once again confirming our experimental hypothesis. Furthermore, as in Experiment 2, Fact Number and Ascription Target did not interact, $F(6, 4296) = 1.16$, $p = .327$. Finally, in analyses collapsing across Fact Number, Ascription Target did not interact with either age, $F(1, 716) = 2.26$, $p = .13$, or gender, $F(1, 716) = 0.01$, $p = .91$.

Did being called Alex, or having a friend named Alex, moderate truth likelihood estimates? Any conclusions regarding the former are moot, given that only 1.3% of the sample (six in the control condition) reported having that name. In contrast, a substantial minority 17.5% (66 in the control condition) reported being friends with an Alex. Planned contrasts within the control group, however, found no sign of a difference between this minority and the majority, $t(90.19) = .03$, $p = .97$ (respective Ms = 57.21 vs. 57.27), suggesting that the SPOT effect is indeed confined to one's "own" theories. The handful of control participants named Alex did give numerical higher estimates (Ms = 62.07), perhaps justifying future investigation. At all events, being or knowing an Alex did not significantly affect the results of Experiment 3, and, by extension, they were unlikely to have affected the results of previous experiments.

Finally, did any of the four indices of self-enhancement—two traits, and two manifest biases—significantly moderate the SPOT effect? Before addressing the question, we first explored the links among these four indices (Table 1). They formed a coherent positive manifold of small but highly significant correlations. One can infer, for example, that the more narcissistic participants were, and the more unimpeachable they regarded their judgment as being, the more they claimed to recognize fake relative to real words, and the less their evaluations of argument quality accorded with more objective consensus estimates. A plausible interpretation of the positive manifold is that the four indices converge in assessing self-enhancing tendencies.

To determine whether any of the self-enhancement indices significantly moderated the SPOT effect, we conducted separate analyses for each index in which we regressed truth likelihood estimates on (a) a *z*-score version of that index, (b) a z-score version of the experimental manipulation, its conditions coded as +1 and -1; and (c) the multiplicative

product of both these z-scores. Both (a) and (b) were entered into the model first, so that the residual independent impact of (c)—the interaction terms representing the moderation of the SPOT effect—could be tested. This method maximizes statistical power and controls for inessential collinearity among the predictors (Aiken & West, 1991). However, in no case did the relevant interaction term reach significance ($-.05 < \beta < .002$; $-1.31 < t[718] < .94$; all ps $> .19$). Thus, the SPOT effect, although once again replicated, was not predicted by four other indices of self-enhancement, which nonetheless correlated with one another. One explanation is simply that the SPOT effect is small, thereby curtailing the magnitude of correlations it displays. The effect size in this study was lower than previous studies, and the larger sample size may not have been sufficient to offset it. Alternatively, the SPOT effect may be a form of self-enhancement that is different in character to others, perhaps being more basic or associative in nature (Gawronski et al., 2007; Gawronski & Bodenhausen, 2011). Our null effect here is in keeping with research that has also failed to find expected correlations between different indices of cognitive bias (Oyer, Gillespie, Issah, & Fasko, 2012).

General Discussion

In this article, we demonstrated the existence of a new bias: the SPOT effect. We showed experimentally that, even where theory ownership is established in the most minimal of ways— simply by asking participants to *imagine that a theory is theirs* rather than someone else's or nobody's in particular—participants are inclined to increase their estimates of the likelihood that that theory is true, as they consider successive items of evidence bearing on the truth of that theory. As such, the SPOT effect is simultaneously a type of confirmation bias (Nickerson, 1998), a mere ownership effect (Beggan, 1992), and an instance of self-enhancement (Sedikides & Gregg, 2008).

We also demonstrated that the SPOT effect occurs regardless of gender and age (Experiments 1, 2, and 3), and that it reflects a pro-self as opposed to anti-other bias (Experiment 2). However, we did not establish the precise mechanism underlying it. Nevertheless, that the SPOT effect did not correlate with several indices of self-enhancement, which are intercorrelated (Experiment 3) suggests it may be distinctively associative in character (Gawronski et al., 2007). Future research could profitably explore whether the

SPOT effect correlates with susceptibility to self-enhancement effects of a more associative sort, such as the name letter task (Hoorens, 2014) and whether it sensibly waxes and wanes in response to self-enhancement manipulations that are indicative of motivational underpinnings (cf. Jones, Pelham, Mirenberg, & Hetts, 2002).

How significant is the SPOT effect? The effect sizes we obtained across three experiments ranged from $d = .23$ to $d = .34$, which places them, by convention, into the "small" category (Cohen, 1988). This suggests that, in terms of practical potency, the SPOT effect may be modest: compared to other antecedents, it may not augment greatly people's propensity to be partisan in evaluating the truth of their theories. Indeed, we engineered the conditions that would maximize the likelihood of obtaining the SPOT rather than faithfully mimicking naturalistic settings. For example, we used a theory that was hypothetical and otherworldly, rather than real and relevant, to minimize the possibility that any prior opinions held by participants would constrain their estimates of the likelihood that that theory was true. Furthermore, the manner in which people come to hold theories in everyday life is unlikely to be, as in our manipulation, by arbitrary ascription; rather, it is usually by more or less elaborate forms of ratiocination (Wegener & Carlston, 2005) or as a result of everyday argumentation (Mercier & Sperber, 2011).

However, we would argue that the significance of the SPOT effect does not reside in its practical potency, nor even in its ultimate generality, but rather in its theoretical implications (Mook, 1983). The goal of the current research was to test for the *existence* of a psychological effect which might, on some initially reasonable assumptions, be deemed unlikely a priori, such that its reliable demonstration would then imply something *surprising* about how the mind works. In particular, one might plausibly argue that so trivial a manipulation as drawing a nominal link between the self and a theory should *not* make people inclined to regard that theory as true on the initially reasonable assumption that their rationality was even moderately robust to self-enhancing biases. However, that drawing such a nominal link *does* indeed have this effect only underlines how exquisitely sensitive to self-enhancing biases the human mind actually is (Sedikides & Gregg, 2008; Sedikides & Strube, 1997). Consider: it might well have been the that no self-enhancing partisanship could

emerge unless people had first become personally committed to their theory—for example, by publicly declaring their support for it (cf. Aronson, Fried, & Stone, 1991)—such that a modicum cognitive dissonance would be required to engender partisan bias. Our findings, however, suggest that no such prior commitment (or any other factor) is necessary: even minimally connecting the self to a theory suffices to induce a spontaneous partisanship towards it (cf. Greenwald, Pickrell, & Farnham, 2002).

The same point applies to many cognate effects we reviewed in the introduction. Consider, for instance, the minimal group discrimination effect, and its associated paradigm (Tajfel et al., 1971). Tajfel and colleagues originally devised that paradigm as a baseline to which they expected to have to add further nuances before any group-based discrimination would be observed. They discovered, however, that mere categorization into separate groups was sufficient to make people differentially allocate rewards to fellow group members. This accidental finding underlined the relevance of identity-related categorizations, and spawned decades of research clarifying the precise mechanism underlying the effect as well as two major theories of social identity (Hewstone et al., 2002). That group categorizations are rarely arbitrarily assigned in everyday life, or that the discrimination observed in the paradigm is limited in severity (e.g., does not characterize the allocation of penalties: Mummendey et al., 1992), is secondary: the theoretical significance of the minimal group discrimination effect holds above and beyond its practical potency or ultimate generality.

In closing, the significance of minimal effects, such as the SPOT effect, is that they are phenomena which require surprising little to bring about (e.g., mere ascription of ownership). Also, that they are brought about at all suggests the presence of profound determinants (e.g., self-enhancement motivation) that have considerable theoretical relevance.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: London, Sage.

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M.D. Alicke, D.A. Dunning, & J. I. Krueger, (Eds.), *The self in social judgment*. New York, NY: Psychology Press.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, *68*, 804–825.

Alicke, M. D., Vredenburg, D. S., Hiatt, M., & Govorun, O. (2001). The better than myself effect. *Motivation and Emotion*, *25*, 7–22.

Alicke, M. D., & Sedikides, C. (2009). Self-enhancement and self-protection: What they are and what they do. *European Review of Social Psychology*, *20*, 1-48.

Alicke, M. D., & Sedikides, C. (2011). *Handbook of self-enhancement and self-protection*. New York, NY: Guilford Press.

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, *40*, 440-450.

Aron, A., Mclaughlin-Volpe, T., Mashek, D., Lewandowski, G., Wright, S. C., & Aron, E. (2005). Including others in the self. *European Review of Social Psychology*, *14*, 101-132.

Aronson, E., Fried, C. B., & Stone, J. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health*, *81*, 1636-1637.

Beggan, J. K. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of Personality and Social Psychology*, *62*, 229–237.

Bosson, J. K., Lakey, C. E., Campbell, W. K., Zeigler-Hill, V., Jordan, C. H., Kernis, M. H. (2008). Untangling the links between narcissism and self-esteem: A theoretical and empirical review. *Social and Personality Psychology Compass*, *2*, 1415–1439.

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter. *Personality and Social Psychology Bulletin*, *38*, 209–219.

Buunk, P. (2001). Perceived superiority of one's own relationship and perceived prevalence of happy and unhappy relationships. *British Journal of Social Psychology*, *40*, 565–574.

Carmon, Z., & Ariely, D. (2000). Focusing on the forgone: How value can appear so different to buyers and sellers. *Journal of Consumer Research*, *27*, 360–370.

Chatterjee, P., Irmak, C., & Rose, R. L. (2013). The endowment effect as self-enhancement in response to threat. *Journal of Consumer Research*, *40*, 460-476.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

College Board (1976–1977). *Student descriptive questionnaire*. Princeton, NJ: Educational Testing Service.

Cross, K. (1977). Not can but will college teachers be improved? *New Directions for Higher Education*, 17, 1–15.

Ditto, P. H., Munro, G. D., Apanovich, A. M., Scepansky, J. A., & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin*, *29*, 1120-1132.

Dunning, D. (2014). Motivational theories. In B. Gawronski & G. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 108-131). New York, NY: Guilford Press.

Feys, J. (1991). Briefly induced belongingness to self and preference. *European Journal of Social Psychology*, *21*, 547-552.

Finch, J. F., & Cialdini, R. B. (1989). Another indirect tactic of (self-)image management: Boosting. *Personality and Social Psychology Bulletin*, *15*, 222-232.

Gawronski, B., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59-127.

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*, 221-232.

Gebauer, J. E., Göritz, A. S., Hofmann, W., & Sedikides, C. (2012). Self-love or other-love? Explicit other-preference but implicit self-preference. *PLoS One*, *7*, e41789

Gramzow, R. H., & Gaertner, L. (2005). Self-esteem and favoritism toward novel in-groups: The self as an evaluative base. *Journal of Personality and Social Psychology*, *88*, 801-815.

Green, J. D., Sedikides, C., & Gregg, A. P. (2008). Forgotten but not gone: The recall and recognition of self-threatening memories. *Journal of Experimental Social Psychology, 44*, 547-561.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4-27.

Greenwald, A. G., Pickrell, J. E., & Farnham, S. D. (2002). Implicit partisanship: Taking sides for no reason. *Journal of Personality and Social Psychology*, *83*, 367-379. doi:10.1037/0022-3514.83.2.367.

Gregg, A. P., Mahadevan, N., & Sedikides, C. (2015*). A revised version of the Argument Evaluation Test*. Manuscript in preparation, University of Southampton.

Gregg, A. P., & Sedikides, C. (2014). Self-enhancement. In A. C. Michalos (Ed.), *Encyclopedia of Quality of life and well-being research* (pp. 5769-5771). Dordrecht, the Netherlands: Springer.

Gregg, A. P., Seibt, B., & Banaji, M. A. (2006). Easier done than undone: Asymmetry in the malleability of automatic preferences. *Journal of Personality and Social Psychology*, *90*, 1-20.

Guenther, C. L., & Alicke, M. D. (2010). Deconstructing the better than average effect. *Journal of Personality and Social Psychology*, *99*, 755-770.

Hamamura, T., Heine, S. J., & Takemoto, T. (2007). Why the Better-than-Average Effect is a worse-than-average measure of self-enhancement. An investigation of conflicting findings from studies of East Asian self-evaluations. *Motivation and Emotion*, *31*, 247-259.

Helweg-Larsen, M., & Shepperd, J. A. (2001). Do moderators of the optimistic bias affect personal or target risk estimates? A review of the literature. *Personality and Social Psychology Review*, *51*, 74–95.

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575-604.

Hoorens, V. (2014). What's really in a Name-Letter Effect? Name-letter preferences as indirect measures of self-esteem. *European Review of Social Psychology*, *25*, 228-262.

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004).How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology*, 87, 665-683.

Jones, J. T., Pelham, B. W., Mirenberg, M. C., & Hetts, J. J. (2002). Name letter preferences are not merely mere exposure: Implicit egotism as self-regulation. *Journal of Experimental Social Psychology*, *38*, 170–177.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, *98*, 1325-1348.

Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin*, *23*, 736-742.

Klayman, J., & Ha, Y-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211-228.

Koole, S. L., & Pelham, B. W. (2003). On the nature of implicit self-esteem: The case of the name letter effect. In S. Spencer, S. Fein, & M. P. Zanna (Eds.), *Motivated social perception: The Ontario Symposium*. (pp. 93-116). Hillsdale, NJ: Erlbaum.

Kruger, J. (1999). Lake Wobegon be gone! The below-average effect and the egocentric nature of comparative ability judgements. *Journal of Personality and Social Psychology*, *77*, 221-232.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480-498.

Lee, S., Gregg, A. P., & Park, S. (2013). The person in the purchase: Narcissistic consumers prefer products that distinguish them. *Journal of Personality and Social Psychology*, 1*05*, 335-352.

Mandel, D. R. (2002). Beyond mere ownership: Transaction demand as a moderator of the endowment effect. *Organizational Behavior and Human Decision Processes*, *88*, 737-747.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–111.

Mook, D. G. (1983). In defense of external validity. *American Psychologist*, *38*, 379–387.

Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes*, *102*, 42–58.

Morewedge, C. K., Shu, L. L., Gilbert, D. T., & Wilson, T. D. (2009). Bad riddance or good rubbish? Ownership and not loss aversion causes the endowment effect. *Journal of Experimental Social Psychology*, *45*, 947-951.

Mummendey, A., Simon, B., Dietze, C., Grunert., M., Haeger, G., Kessler, S., Lettgen, S., & Schaferhoff, S. (1992). Categorization is not enough: Intergroup discrimination in negative outcome allocation. *Journal of Experimental Social Psychology*, *28*, 125-144.

Nesselroade, K. P.,Jr., Beggan, J. K., & Allison, S. T. (1999). Possession enhancement in an interpersonal context: An extension of the mere ownership effect. *Psychology & Marketing*, *16*, 21–34.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.

Nuttin, M. J. (1987). Affective consequences of mere ownership: The name letter effect in twelve European languages. *European Journal of Social Psychology*, *17*, 381-402.

Otten, S. (2004). Self-anchoring as predictor of in-group favoritism: Is it applicable to real group contexts? *Cahier Psychologie Sociale*, *22*, 427-443.

Otten, S. (2005). The in-group as part of the self: Reconsidering the link between social categorization, in-group favoritism and the self-concept. In A. Alicke, D. Dunning, & J. Krueger (Eds.), *The self in social perception* (pp. 241-265). Philadelphia: Psychology Press.

Oyer, B., Gillespie, M., Issah, M., & Fasko, D. (2012). The role of personality in argument evaluation. *Inquiry: Critical Thinking across the Discipline*, *27*, 40-49.

Pahl, S. & Eiser, J. R. (2005) Valence, comparison focus, and self-positivity biases: Does it matter whether people judge positive or negative traits? *Experimental Psychology*, *52*, 2005, 303-309.

Paulhus, D. L. (1998). *Manual for Balanced Inventory of Desirable Responding* (BIDR-7). Toronto: Multi-Health Systems.

Paulhus, D. L., & Harms, P. D. (2004). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*, 890–904.

Presson, P. K., & Benassi, V. A. (1996). Illusion of control: A meta-analytic review. *Journal of Social Behavior and Personality*, *11*, 493-510.

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, *43*, 565-578.

Roth, J., & Steffens, M.C. (2014). When I becomes we: Associative self-anchoring drives implicit intergroup bias in minimal groups. *Social Psychology*, *45*, 253-264.

Schmitt, D.P., & Allik, J. (2005). Simultaneous administration of the Rosenberg Self-Esteem Scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, *8*9, 623-642

Searle, J. R. (1995). *The construction of social reality*. New York, NY: Free Press.

Sedikides, C., & Alicke, M. D. (2012). Self-enhancement and self-protection motives. In R. M. Ryan (Ed.), *Oxford handbook of motivation* (pp. 303-322). New York, NY: Oxford University Press.

Sedikides, C., Gaertner, L., Luke, M. A., O'Mara, E. M., & Gebauer, J. (2013). A three-tier hierarchy of motivational self-potency: Individual self, relational self, collective self. *Advances in Experimental Social Psychology*, *48*, 235-296.Sedikides, C., Gaertner, L., & Cai, H. (2015). On the panculturality of self-enhancement and self-protection motivation: The case for the universality of self-esteem. In A. J. Elliot (Ed.), *Advances in Motivation Science* (Vol. 2, pp. 185-241). San Diego, CA: Academic Press. doi:10.1016/bs.adms.2015.04.002

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology*, *84*, 60–79.

Sedikides, C., & Green, J. D. (2000). On the self-protective nature of inconsistency/negativity management: Using the person memory paradigm to examine self-referent memory. *Journal of Personality and Social Psychology*, *79*, 906-922.

Sedikides, C., & Green, J. D. (2009). Memory as a self-protective mechanism. *Social and Personality Psychology Compass*, *3*, 1055-1068.

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, *3*, 102-116.

Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. (2014). Behind bars but above the bar: Prisoners consider themselves more prosocial than non-prisoners. *British Journal of Social Psychology*, *53*, 396-403.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology, 29*, 209-269.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology*, *89*, 342-357.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, *22*, 259–264.

Svenson, O. (1981). Are we less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*, 143–151.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*, 755–769.

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149-178.

Taylor, S. E., & Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193-210.

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, *106*, 1039–1061.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–28.

Wegener, D. T., & Carlston, D. E. (2005). Cognitive processes in attitude formation and change. In D. Albarracin, B. Johnson, & M. Zanna (Eds.), *The handbook of attitudes* (pp. 493-542). Mahwah, NJ: Erlbaum.

Wenger, A., & Fowers, B. J. (2008). Positive illusions in parenting: Every child is above average. *Journal of Applied Social Psychology*, *38*, 611–634.

Figure Captions

Figure 1

Experiment 1: Participants' Estimates of the Likelihood of the Theory Being True as a Function of Fact Number and Theory Ascription

Figure 2

Experiment 2: Participants' Estimates of the Likelihood of the Theory Being True as a Function of Fact Number and Theory Ascription

Figure 3

Experiment 3: Participants' Estimates of the Likelihood of the Theory Being True as a Function of Fact Number and Theory Ascription

Table 1

Experiment 3: Intercorrelations between the Four Indices of Self-Enhancement

	NPI	SDE	OCL	BAE
NPI	1			
SDE	.34***	1		
OCT	.11*	.12*	1	
BAE	.14**	11*	.13*	1

Note. N = 703-717. NPI = Narcissistic Personality Inventory, 16-item version. SDE = Self-Deceptive Enhancement (a subset of seven judgement-relevant items from the Biased Inventory of Desirable Responding). OCL = Overclaiming (the ratio of fake to real words that participants claimed to recognize). BAE = Biased Argument Evaluation (the extent to which participants' judgments of argument quality diverged from consensus estimates).

* $p < .005$. ** $p < .0005$. *** $p < .00005$