New Ideas in Psychology 31 (2013) 73–86

Contents lists available at SciVerse ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

Comparisons in research and reasoning: Toward an integrative theory of social induction

Joachim I. Krueger*, David Freestone, Mika L.M. MacInnis

Brown University, USA

Keywords:
Theoretical explanations
Experimentation
Rationality

A B S T R A C T

Strong scientific theories give coherence to a body of research findings, make precise predictions about key phenomena, and guide the search for new discoveries. In social psychology, some contemporary theories fall short of this ideal. Mini-theories are prevalent (cf. Van Lange, Higgins, & Kruglanski, 2011), many predictions are merely directional (*like this one!*) and theorizing *post-hoc*. Guided by experimental reasoning, many researchers emphasize—and *reify*—empirical differences. Taking the experimental method as an epistemological gold standard, they regard comparative thinking as a criterion of rational thinking. Using examples from social judgment and decision making, we show how comparative reasoning can constrain theoretical development and bias assessments of human rationality. To encourage movement toward stronger theory, we describe a model of inductive reasoning in social contexts.

1. Introduction

> *Comparison* is essential in research and should be omnipresent.
>
> ~Ellsworth and Gonzalez (2003, p. 25; *italics* in original)

> Every man bears the whole stamp of the human condition.
>
> ~Montaigne (in "Of Repentance")

The sciences seek to provide insights into nature where simple observation fails. To succeed, any science needs an epistemology and a set of methods and procedures that guide the search for knowledge (Lakatos, 1978). Psychological science is in a unique position because one of its tasks is to study how ordinary people draw inferences from observations. To us, it is not surprising that psychological science has approached the study of everyday inference from the point of view of its own epistemological and methodological commitments. For scientists to expect that people think as they themselves do is an intriguing instance of social projection (Krueger, 1998). A crucial feature of this process is that it tends to become normative. That is, scientists come to demand that ordinary people think like scientists do. Influential psychologists have proposed that people can be regarded as rational inasmuch as they master the logic of deduction (Inhelder & Piaget, 1958; Wason, 1960), or the calculus of probability (Nisbett & Ross, 1980; Tversky & Kahneman, 1974).

The present article is concerned with the interplay between the general scientific enterprise of gathering knowledge and the implications of this enterprise for research on everyday rationality. We focus on the logic of experimentation as a dominant feature of the scientific method in psychology. Coupled with standard methods of null hypothesis significance testing (NHST), the experimental paradigm leads some researchers to construe human rationality as reasoning by comparisons. We

* Corresponding author. Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Box 1821, 190 Thayer St., Providence, RI 02912, USA. Tel.: +1 401 863 2503.
E-mail address: Joachim_Krueger@Brown.edu (J.I. Krueger).
URL: http://research.brown.edu/research/profile.php?id=10378

0732-118X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.newideapsych.2012.11.002

identify weaknesses in contemporary research practice and explore their implications for the study of rationality. The heart of our argument is that with improvements in theory development and hypothesis testing, the study of human rationality will also benefit.

We focus our inquiry on the social psychology of judgment and decision making. This subdiscipline of psychology relies heavily on experimental design to formulate research questions, and on NHST to evaluate people's answers. Specificity is seen as a virtue and there are few attempts to integrate sets of phenomena into comprehensive theoretical frameworks (Edwards & Berry, 2010; Higgins, 2004; Kenrick, 1994). Concurrent with these epistemic and methodologicial challenges, the assumption that ordinary people should reason by comparison like experimentalists do lies close to the field's axiomatic core (Gigerenzer, 1991). Therefore, any critique of experimental or data analytic methods has implications for conceptions of rationality. We pursue this theme as the central objective in this article, noting that the linkage between a critical assessment of methods of design and analysis on the one hand, and the criteria used to evaluate human rationality on the other, has gone largely unexamined in the scientific literature.

Our first concern is that the search for significant differences opens the door to *post-hoc* theorizing (Kerr, 1998). The discovery of differences constrains the depth of theory if these differences are *reified*, that is, if they are treated as things instead of relationships, and when corresponding latent psychological constructs are postulated to explain these differences. Ellsworth and Gonzalez (2003) are clear about the need to distinguish between the two. While they insist that comparisons are essential to research (see epigraph), they also caution not to "discuss effect sizes in a manner that implies a deep fundamental relation" (p. 38). In a particular instance of research, it may be difficult to decide whether the fallacy of reification has occurred. The most obvious case is one that involves a tautology, that is, when the statistical effect and the process that presumably caused it are called by the same name and are not measured independently. Higgins (2004), for example, suggested just this happened in the case of the most famous phenomenon in the literature on social perception. He writes that:

> "social psychologists represented the special phenomenon as being *about* people underestimating the impact of situational forces and then they inferred a special "fundamental attribution error" or "correspondence bias" as the source of the phenomenon by inferring that the source of the special phenomenon was a special mechanism" (p. 309).

The strategy of finding and reifying differences contrasts with the ability of strong theories to parameterize underlying psychological processes and to make precise quantitative predictions at the level of perception or behavior (Kruglanski & Gigerenzer, 2011). As Busemeyer and Wang (2000, p. 171) put it, "accurate *a priori* predictions to new conditions are the hallmark of a good scientific theory."

The experimental, difference-seeking, epistemology has become a metaphor of mind. There is a direct – and perhaps projective – link between how psychologists think about research design and how they think about human reasoning (Gigerenzer, 1991). As experimentation is often considered the strongest and most rational way of doing research, it is not surprising that experimenters themselves have modeled human rationality after the experimental ideal. This argument is intuitively appealing. Humans attempt to learn about nature, but nature reveals her secrets only reluctantly. Controlled experiments are a powerful method to pose questions in such a way that compel nature to yield relevant data. If the epistemology of controlled laboratory experiments and ordinary perception are essentially the same, it seems clear that what people need to do in their everyday thinking is to approximate—if not replicate—the strategies involved in experimentation. Doing so will make their reasoning rational. If the key to experimentation is the controlled variation of conditions and the statistical comparison of results between conditions, then the hallmark of rationality should also be variation and comparison. Indeed, people can improve their judgments, or at least recognize, the limitations of what they think they know by considering the opposite, counterfactuals, or non-sampled outcomes (Hastie & Dawes, 2010).

Despite these undeniable virtues of experimental reasoning, we suggest that the attempt to model everyday rationality on scientific method bears certain risks. We show that the mandate of looking for differences often fails to tell us *which* differences are relevant. With the example of research on self-evaluation and intergroup perception, we show that individuals (and the psychologists who study them) engage in a variety of comparisons resulting in contradictory inferences regarding human rationality.

In the final section of this article, we return to the question of how theories that are both strong and parsimonious overcome some of the barriers raised by the conventional strategy of seeking and isolating differences, reifying them, and passing judgment on the ordinary perceiver. To illustrate an opportunity for progress, we describe a theory of social perception that integrates a variety of phenomena that are traditionally studied in isolation, makes precise point-predictions, and does not hasten inferences about social perceivers' (ir)rationality.

1.1. Experimentation

Since Wilhelm Wundt (1874) worked to emancipate psychology from philosophy, the laboratory experiment has been the backbone of psychological research (Runkel & McGrath, 1974). There is an elegance to a well-designed experiment. At its best, an experiment is a controlled intervention that harnesses Mill's (1872/1973) method of difference. A difference between treatment and control conditions forces nature to support one causal claim over another. At worst, this esthetic obscures the question of why the experiment was conducted in the first place. Today, the goal of most psychological experiments is to reject an undesired null hypothesis (Gallistel, 2009; Krueger, 2001). Most null hypotheses express the expectation that the true effect size is exactly zero. A theory seeking to corroborate a substantive claim by rejecting a null hypothesis is weak (Meehl, 1978). It only makes a directional claim. A theory

that is corroborated by the rejection of a null hypothesis from either direction is essentially worthless.

Reviewers and meta-analysts bemoan the inflation of trivial, uninterpretable, or irreplicable findings (Bakker & Wicherts, 2011; Fanelli, 2010, 2012; Fiedler, 2011; Ioannidis, 2005) and some demand that the bias against the null hypothesis be reduced by making significance testing more conservative (Simmons, Nelson, & Simonsohn, 2011). Others, however, claim that a one-sided attack on α-errors will do nothing to improve the development of generative and integrative theories (Fiedler, Kutzner, & Krueger, 2012).

To appreciate the scientific hollowness of atheoretical significance testing, consider Bem's (2011) study of precognition and backward causation. From statistically significant results, Bem concludes that random events occurring in the future can causally determine human choices in the present. NHST, however, cannot speak to the theoretical question of why this should be so. Instead of offering a positive process-based account, Bem uses the absence of a known process to make what he takes to be an *a fortiori* argument for the existence of a nonstandard process – the identity of which remains to be determined. Until that time, Bem *reifies* his finding as "psi," conflating the statistical difference of the result with the presumed process that produced it.

Bem's inference strategy seeks to get two inferences for the price of one. This strategy is not limited to paranormal work. Consider the influential research program on automaticity. In this paradigm, subtle manipulations, such as subliminal primes, are shown to affect some behavior of interest (e.g., mimicry). With the rejection of the null hypothesis of chance variation, the inference is that the causal role of automatic priming has been verified, and that the folk idea that conscious processes control behavior has been falsified. This inference is reached in a "subtractive manner" (Bargh, 2007, p. 557), which requires the assumption that automatic and controlled processes are related hydraulically (i.e., that they can be represented in a zero-sum model). There is, however, an asymmetry in the method of subtraction. It does not work when the null hypothesis is *not* rejected. In that case, subtraction would imply that the behavior was consciously controlled, but NHST suggests that the behavior was random. If the hydraulic principle is invoked only when there is significant evidence for automaticity, there can be no positive evidence for control.[1]

Experimental designs, and the statistical analyses associated with them, vary in the depth and the coherence of the theoretical inferences they afford. Many of the widely accepted theories in psychology remain weak in the sense that they fail to make specific point predictions that have a record of surviving attempts at falsification (Popper, 1963). In the social psychology of judgment and decision making, this weakness is rather obvious, a state of affairs that is, in part, a historical accident of Tversky and Kahneman's (1974) decision to show the power of heuristic thinking not in parametric experiments but in a series of vivid, memorable, and story-like demonstrations. This method only allows ordinal predictions.

1.2. Experimental thinking

When experimental thinking is the ideal of scientific thinking, it is natural to ask whether ordinary cognition measures up. If even the simplest experiment requires a comparison, then rational or scientific reasoning must also proceed by making critical comparisons. Dawes (2001) suggested that humans systematically fail to think rationally. Instead, they rely on non-comparative or associative reasoning. Associative reasoning tracks linkages among stimuli or impressions and it is, according to philosopher John Locke (1690/1970), the bedrock of all reasoning. Hastie and Dawes (2010) argue that associative thinking is not enough because depends on individual salient stimuli and not on comparisons.

But then again, merely noting that a comparison has been made is not enough to establish rationality. There are two more specific criteria. One is that judgments not contradict each other (coherence), and the other is that judgments not depart from an objectively correct response (correspondence). As Kahneman and Tversky (1982, p. 124) put it, "the presence of an error of judgment is demonstrated by comparing people's responses either to an established fact […] or to an accepted rule of arithmetic, logic or statistics."

By these definitions, people reason rationally inasmuch as they minimize, or even eliminate, differences between judgments or between judgment and criterion. This raises the conundrum that *people are asked to think in terms of differences in order to avoid differences in their judgments*. In terms of statistical evaluation, rationality becomes a null hypothesis of no difference, which means that a research methodology that is focused on the detection of differences can only produce positive findings of human irrationality (Krueger & Funder, 2004). From this perspective it is not surprising that the catalog of biases and errors has grown steadily (Gilovich & Griffin, 2010; Herbert, 2010).

While it is true that Dawes (2001) struck an important blow for rationality by demanding that people look beyond the information given in a certain situation, it is also true that this standard is often hard to meet. Ordinary people cannot run control conditions to observe how their own behavior would change. Ordinary choice between behavioral options means that as one option is selected another one is foregone. When they cannot directly observe and compare alternative outcomes, people can only simulate "events and relationships that are not salient and explicit in [their] experience" (Hastie & Dawes, 2010, p. 114). But even if people simulate unknown conditions, they cannot account for the *unknown unknowns*, to use former U.S. Secretary of Defense Donald Rumsfeld's memorable phrase (cited in Keeley, 2007). Unknown unknowns are experiences they do not even realize they are missing. Phil, when asked how he was doing, demanded "Compared to what?"[2]

[1] Alternative methods, which generate separate estimates for the contributions of automatic and controlled processes, are available (Jacoby, 1991; Krueger, 2009; Payne & Iannuzzi, 2012).

[2] We thank Josh Klayman for bringing this joke to our attention.

The heuristic of anchoring and (insufficient) adjustment illustrates the difficulty of performing and integrating multiple comparisons. Anchoring is defined as the tendency to overweigh initial estimates or values, even when these are clearly irrelevant (Frederick & Mochon, 2012; Tversky & Kahneman, 1974). Hastie and Dawes (2010) consider anchoring the most fundamental heuristic because it has been applied to the greatest variety of judgment and decision-making tasks.

Imagine your task is to estimate the population of Greenland. Before you do, however, you are asked if the number is greater than 10 million. Alternatively, you are first asked if the number is smaller than 1000. Your answer to both questions is "No," which confirms that both extreme numbers are irrelevant for the estimation task. Next, you provide your estimate of the population and your number will likely by higher if you first rejected a high anchor than if you rejected a low anchor.

The difference between the two estimates represents the size of the bias, but this difference is itself a composite of two differences. The downward adjustment from the high anchor produces one difference, and the upward adjustment from the low anchor produces another. The anchoring effect can be written as:

$$(HighAnchor - LowAnchor) - [(HighAnchor - HighEstimate) + (LowEstimate - LowAnchor)].$$

This decomposition of a simple reified difference reveals that what is described as the anchoring effect depends on four distinct psychological processes, two of which are associative (respect for the two anchors), and two that are comparative (adjustments away from them). Rationality demands that the impact of the two associative processes be nullified by the two adjustments. This goal is reached only if the sum of the two adjustments is equal to the difference between the two anchors. This example suggests that the pursuit of rationality through the construction of "relationships that are not salient and explicit in our experience" (Hastie & Dawes, 2010, p. 114) can be rather difficult.

How is this problem supposed to be solved? To say "Just ignore the anchor," is not helpful because people already try to do this. They recognize the anchor's irrelevance. An alternative strategy is for an individual judge to also generate a low anchor (if only given a high anchor, and *vice versa*), to simulate what the two final, post-adjustment, estimates would be, and then to split the difference between the two. There is a chance that people can do this sort of dialectical bootstrapping if asked (Herzog & Hertwig, 2009), but there is no evidence that they will do it spontaneously.

To review, we note that experimental psychology construes the discovery of knowledge as a task of detecting differences. We also observe that there is a common belief that rational reasoning ought to be experimental in nature (Dawes, 2001; Nisbett & Ross, 1980). Failure to meet this demand is seen as the footprint of associative, heuristic, or automatic reasoning.

The example of the anchoring effect suggests that ordinary perceivers may be hard pressed to replicate many of the experimental variations that can be devised for a laboratory study. The life of an ordinary person is, as it were, an experiment without a control group. Even granting that ordinary people cannot conduct true experiments, the demand for comparative reasoning remains. Several prominent theories of social perception ask that people notice and encode differences between relevant and accessible bits of information. Their judgments and decisions should then depend on these differences rather than on the individual components of these differences.

In the next section, we explore the intersection between the difference-seeking research strategy and its implications for rationality in everyday judgment in the area of self-evaluation and intergroup perception. We show that this approach enables errors of reification and limits theoretical development.

2. Comparative reasoning in social context

2.1. Self-perception

The experimental mindset demands that the "Compared to what?" question be asked for a judgment not only to be rational, but even to be comprehensible (Gilovich, 1997). Applying this idea to self-perception, Karniol (2003) proposed that the self-concept rests on implicit comparisons. A person first mentally draws up a profile of attributes characteristic of the average person, or *protocenter*. Then the person compares him- or herself with this protocenter and encodes only the differences in the self-concept. This process amounts to a correction for the attribute base rates.[3] It captures the logic of any method that residualizes judgment, such as experimental comparison, multiple regression, or the logic of conversation (Grice, 1975). Each of these methods is designed to remove redundant information and to retain what is diagnostic.

The comparative method facilitates some types of judgment. Indeed, it is necessary for classification and recognition. A pouch is critical for an animal to be recognized as a marsupial, whereas fur is not. In contrast, there is no evident need for people to construct self-images as if they were trying to pick themselves out of a line-up. It does not seem irrational for a person to see herself as friendly, knowing that many others also see themselves as friendly.

Empirical research suggests that people naturally describe themselves with a mix of shared and distinctive attributes. Shared attributes are popular, especially when they are positive (Meehl, 1956). Likewise, the process of social projection promotes the perception of self-other similarities. An idealistic person may believe in her own fundamental goodness, and also attribute goodness to people in general (Krueger, Acevedo, & Robbins, 2006). Including shared attributes in the self-image does not

[3] Classic attribution theory takes essentially the same approach to inferences about personality. A behavior is seen as caused by the person if consensus was low, that is, if few others behave this way (Kelley, 1967).

threaten a person's sense of uniqueness. Uniqueness still lies in the personal *pattern* of those attributes.

In theory, all of a person's attributes can be those of the majority, that is, the protocenter. Yet, profile matching becomes less probable, and the person more unique, as the list of statistically independent attributes gets longer. For example, if there is a majority of 80% for each trait on the list, the probability of a perfect match between person and protocenter is .102 for a list of 10 traits, and 2.04^{-10} for a list of 100 traits. According to protocenter theory, a person matching the majority profile would have no self-concept at all. That person would be a *Mann ohne Eigenschaften* (man without qualities) to borrow Musil's (1930) famous phrase. A psychologically plausible alternative is that a person who matches the protocenter on each attribute is a *prototype*, the person that is the most typical exemplar of the group or category. The concept of prototypes is prominent in cognitive and social psychology, and it has deep roots. In his pioneering research on social measurement, Quetelet (1835/1968) conceived of human populations as distributions of individuals around a true center, where each distance between individual and center is a measure of "error" rather than "truth." From this point of view, a person who matches the profile of averages is an ideal.

Protocenter theory implies that if comparisons are made, the referent of the comparison should come to mind first. Therefore, information about the protocenter should be more accessible than information about the self, because the latter can only be constructed relative to the former. In contrast to this view, response-time data suggest that under most conditions judgments about the self are more accessible than judgments about others (Krueger, 2003). There is, however, an interesting exception to this pattern. Under certain conditions (e.g., threat, minority status), people access group-related attributes faster than they access their own personal attributes. As a result, they shape their self-concepts to match the group stereotype (Latrofa, Vaes, Cadinu, & Carnagh, 2010). Yet, this self-stereotyping is also inconsistent with protocenter theory because it comprises a difference-erasing or assimilative process instead of a difference-enhancing or contrastive one.

Another problem of protocenter theory is conceptual rather than empirical. If people attempted to be rational by making protocenter – self comparisons, they would first have to answer the question "Which protocenter?" Consider Imre Lakatos, the noted historian and philosopher of science. Is his protocenter the average philosopher, the average Hungarian, or perhaps the overall average man? Without *a priori* constraints, *any* individual can be perceived against the background of innumerable protocenters as each individual belongs to innumerable groups. With multiple possible protocenters, there are many potential self-protocenter comparisons, and thus many resulting self-concepts.

Protocenter theory also faces a problem of internal coherence. Suppose each member of a group is informed that the protocentric trait profile ranges from a low number of 0 to a high number of 100, where each number is the average self-judgment. If individuals then encode the difference between their self-perception and the protocenter value on each trait, the result will be a uniform average of 50 across traits. This result has an interesting implication. For a protocenter to maintain variability over traits, it is necessary that individual perceivers do *not* reason in the way assumed by protocenter theory. What remains is the possibility that the protocenter is an illusory construction. For example, individuals might endorse positive traits because they believe most others do not (see discussion of self-enhancement below).

We have treated protocenter theory as an example of the attempt to understand self-perception as the rational encoding of differences. The differences are regarded by the researchers (and presumably by the ordinary perceiver) as the only attributes that matter. The differences thereby become reified as the person's self-concept. Protocenter theory is silent on potential biases in the process of comparison. Yet, such biases are the focus of much social psychological research. Indeed, the dominant view is that mental comparisons between self and others are subject to contrast or assimilation effects (Mussweiler, 2003). An important variant of this idea is that people reason comparatively in order to construct a positive, if biased, self-concept. In other words, while the process of comparative thinking is upheld as rational, its outcome is not. We next review research on self-enhancement, and note the lack of agreement about *which* comparison is made, by the person or the researcher. This lack of agreement suggests that claims of irrationality are fragile.

The most popular way of measuring self-enhancement is rooted in the theory of social comparison. Festinger (1954) proposed that people are motivated to achieve and maintain a balance between being similar to, and different from, important others (see also Brewer, 1991). The motive of self-enhancement requires the continual operation of processes that compare the self with others. Yet in most research on self-enhancement, there is little attention to who the important others are. Instead, participants are asked to judge themselves in comparison to the average person. Like protocenter theory, self-enhancement theory tends to bypass the question of which population the average is meant to represent. The typical finding is the Better-Than-Average (BTA) effect (Brown, 2012; Guenther & Alicke, 2010).

Measures of the BTA effect are either direct or indirect. The direct measure is obtained by asking participants to rate themselves on a scale ranging from "I am much worse than the average person" to "I am much better than the average person." This method produces a single number per person and attribute. The presumed validity of the measure rests on the assumption that the person computes an absolute self-judgment and an absolute other-judgment and then translates the difference to the comparative scale. If a person did this rationally and consistently, the resulting comparative judgments would be positively correlated with absolute self-judgments and equally strongly, but negatively, correlated with absolute other-judgments.

As the finding of self-enhancement itself questions rationality, it seems doubtful that people weigh the two types of absolute judgment equally. Indeed, research with the indirect measure suggests that they do not. The indirect measure requires participants to make absolute self- and other-judgments separately, with the observed differences indexing self-enhancement. Klar and Giladi (1999) showed

that absolute self-judgments predict comparative judgments far better than absolute other-judgments do (see Chambers & Windschitl, 2004, for a review). This result suggests the presence of a generic salience bias *sensu* Hastie and Dawes (2010), and that people are not rational in Karniol's sense.[4]

Critics of the social-comparison approach argue that an individual's judgment of the average person is irrelevant for the assessment of self-enhancement. To capture self-enhancement, it is necessary to subtract an estimate of what the person "is really like" from the person's self-judgment. Many personality researchers recommend aggregated observer judgments as a reality criterion (Colvin, Block, & Funder, 1995; John & Robins, 1994). Although the "social-reality" approach of conceptualizing self-enhancement as the difference between self-judgment and average observer-judgment is an improvement over the traditional social comparison approach, it continues to reify a difference score as the measure of a within-person psychological trait. The implication of this approach is that a rational person would know how the average observer perceives him or her and endorse their view in his or her self-concept.

The social reality measure of self-enhancement did not displace the social comparison measure. Work with both types of measure continued, and conflicting results accumulated. Most work with the social comparison measure seemed to suggest that self-enhancement yields psychological benefits, whereas most work with the social reality measure suggested that self-enhancement has undesirable social consequences, at least in the long run (cf. Krueger & Wright, 2011). In an attempt to resolve the matter, Kwan, John, Kenny, Bond, and Robins (2004) proposed an index of self-enhancement that combined social comparison with social reality. They suggested that both a person's judgment of the average other and the average observer's judgment of this person be subtracted from the person's self-judgment. The resulting dual-difference score is meant to capture an idiographic interaction term that shows a unique kind of self-enhancement. Its meaning, however, is difficult to interpret. Different underlying data patterns can yield the same result. One individual may make a favorable comparison *with* others, whereas another person may make a less favorable impression *on* others, and both end up with the same overall difference score (Krueger & Wright, 2011). As already seen with the example of the anchoring-and-adjustment heuristic, composite difference scores potentiate the problems of reification. It is difficult to see how an index that is a difference between differences or the sum of differences could reflect a unitary concept. Any value obtained with such a composite index can arise from many different (even opposite) configurations of the constituent variables.

The following conclusions impress themselves on the observer of this field. First, difference scores are reified to represent the self-concept or self-enhancement. Second, different types of difference score are computed to represent the same underlying construct. Third, the same difference scores are interpreted as evidence of rational comparative thinking by some, but as evidence for irrational egocentric thinking by others. Taken together, it appears that a singular pursuit of differences is not in the best interest of the development of strong theories.

2.2. Social perception

The impressions gathered from research on self-perception and self-enhancement are not unique. Research on group perception shows close parallels. Anticipating Karniol's (2003) protocenter theory, McCauley and Stitt (1978) argued that group stereotypes consist of traits whose prevalence in the group departs from the world base rate. This view still dominates current thinking. In the *Handbook of Social Psychology* (Fifth ed.), Yzerbyt and Demoulin (2010, p. 1035) assert that "people consider that traits are stereotypical of a given group when they think that these traits are more likely in this group than in another group or in the general population."[5] The empirical evidence does not support this claim. Mirroring results in self-perception research, the perceived prevalence of a trait in the group predicts judgments of trait typicality, and differences or ratios involving world base rates add nothing (see Krueger, 2008, for a review).

Like Karniol did in the case of the self-concept, McCauley and Stitt regarded comparative reasoning in the group context as rational. Unlike Karniol, however, McCauley and Stitt offered an explicit Bayesian argument. They suggested that a trait T is stereotypical of a group G if it is more common in the group than in the general population, that is, if $p(T|G)/p(T) > 1$. When multiplied with the size of the group, $p(G)$, the diagnostic ratio yields the probability of categorization (if it has a pouch, the animal is probably a marsupial), but it does not capture group stereotypes. Much as individuals ascribe many shared attributes to themselves, so do they regard attributes that are common in a group as stereotypical, even if these attributes do not differentiate the group from other groups. Protocenter theory can also be cast in Bayesian terms, highlighting the similarities between the rationalist approaches to self- and social perception. For the present purposes, suffice it to say that because of their structural similarities, the other points of critique raised with regard to Karniol's protocenter theory also apply to McCauley's diagnosticity theory.

Much like research on irrational self-enhancement is at odds with protocenter theory, so is research on intergroup relations at odds with McCauley's diagnosticity theory. Most advocates of the conventional view that social stereotypes capture perceived group differences assume that perception is biased and irrational. The group-level analog of individual self-enhancement is the

[4] Suppose everyone were motivated to construct a self-rating that is *x*% above the average person's self-judgment. The result would be a beauty contest (Keynes, 1936; Nagel, 1995), which, if people executed their self-enhancing motive faithfully would lead to a ceiling effect. That, however, is empirically not the case.

[5] Self-categorization theory takes a similar stance, claiming that stereotypes spring from the "principle of meta-contrast" between groups (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987).

concept of ingroup favoritism (cf. Dasgupta, 2004). Mirroring the social comparison approach to self-enhancement, most theories of ingroup favoritism assume that people negatively contrast perceptions of outgroups away from perceptions of ingroups (cf. Tajfel & Turner, 1979). The resulting difference in the perceived favorability of the group has become reified in the concept of ingroup favoritism, which is thought to satisfy a basic social need.

There is also a group-level analog of the social-reality approach to self-enhancement. These theories assume that the true measure of ingroup favoritism holds the target group constant while varying the perceiver's group (Bartsch & Judd, 1993; Park & Rothbart, 1982). This view still reifies the empirical difference as the phenomenon of ingroup favoritism, but does not assume a contrastive causal process. As far as we know, no dual difference score in the mold of the Kwan et al. (2004) interaction term has been proposed for ingroup favoritism.

3. Integrative modeling

We have shown that the experimental mindset has two problematic consequences. First, it encourages the reification of observed differences, which in turn facilitates *post-hoc* theorizing. Second, the mindset of trapping differences hinders attempts to explore the nature of human rationality. Humans are called upon to reason like experimentalists (i.e., in terms of differences), yet some of the results of this reasoning are held against them as evidence for irrationality (e.g., self-enhancement or stereotyping). A third problem, to which we alluded in the introduction, is that research findings tend to remain disparate and isolated. In our view, the tendency to reify specific effects, label them, and construct mini-theories to account for them, hinders progress toward more integrative theories.

In the final section of this article, we show that it is possible to build strong theories that are parsimonious in their assumptions and generous in their yield. Such theories account for known phenomena, predict others not yet investigated, and integrate all within a common framework (Guilhardi & Church, 2009). To show how theories of social perception can become stronger and less judgmental with regard to human rationality, we sketch and extend the Inductive Reasoning Model (IRM), which we introduced earlier (DiDonato, Ullrich, & Krueger, 2011; Krueger et al., 2006; Krueger & DiDonato, 2008). Our present sketch shows how the model reconciles apparently contradictory theoretical views and how it generates new research opportunities.

The IRM takes basic empirically grounded phenomena as input, and it derives predictions regarding higher-order phenomena as output. The empirical backdrop of the model is that most individuals

[1] …have positive self-images.

[2] …make self-judgments that are positively correlated with the average self-judgments in the group (analytically true).

[3] …project their own attributes strongly to ingroups.

[4] …project their own attributes only weakly to outgroups.

In any particular data set, each of these findings is expressed by person-specific correlation coefficients that capture the direction and the strength of the association. From this basic input, the IRM derives precise predictions regarding the phenomena of self-enhancement, ingroup favoritism, intergroup differentiation (perceived group differences), and differential accuracy of ingroup and outgroup perceptions (typically with higher accuracy for the former). For each phenomenon, the model generates point-hypotheses that need not be nil hypotheses. Significance testing can thus be used for strong inferences. A significant result signals the incompleteness of the model (if not its falsity; Edwards & Berry, 2010; Fraley & Marks, 2007; Meehl, 1990).

In the present exposition, we illustrate the properties of the model by looking at the phenomena of self-enhancement and ingroup favoritism. A full quantitative treatment will be presented elsewhere (Krueger, Freestone, & Heck, in preparation).

3.1. Self-enhancement

According to the IRM, self-enhancement can be predicted, modeled, and tested as the difference between two correlations: the correlation between self-judgments and judgments of trait desirability, which represents the positivity of the person's self-image, and the correlation between judgments of others (or a group) and desirability judgments, which represents the perceived positivity of the average other. The model does not compel or even suggest a reification of this difference. Indeed, it is hard to see how one would construe the difference between two correlations as an entity.

The IRM not only predicts that most people will show self-enhancement, but also that rationality demands it. The model begins with the empirically well-supported assumption of egocentrism, or the notion that people make self-judgments with relative ease. The accessibility advantage of self-referent information is due, in part, to the fact that individuals tend to have more and better information about themselves than about others and that they are more practiced in retrieving this information (Fiedler, 1996; Windschitl, Rose, Stalkfleet, & Smith, 2008). People then take the inductive step of inferring the traits of the average person from their own traits. This type of inference is a form or social projection (Krueger, 1998), and the model estimates it as the correlation between self-judgments and judgments about the social group. Likewise, the model assumes that most people have positive self-images (Alicke & Sedikides, 2011), a parameter that is also estimated from the empirical data.

To see how self-enhancement can arise from social projection combined with a positive self-image, imagine respondents making self-judgments for a series of traits ranging from the highly undesirable to the highly desirable. Based on the empirical finding that most people endorse

positive traits more strongly than they endorse negative traits, we expect a positive correlation between average self-judgments and trait desirability. The correlation between all individual self-judgments and these averages must be, however, less than perfect because there is variation in the self-judgments for each trait.

How might people approach the task of estimating the group means? The IRM does not assume that people compare themselves with others on a particular trait. Instead, a simple heuristic is sufficient to produce judgments. People may estimate group means by using their own self-judgments and weigh them according to the assumed degree of similarity between the self and the average person. Indeed, the optimal weight is the correlation between self-judgments and group averages over all individuals and traits (Hoch, 1987). The researcher who has access to the data can compute this correlation, but the ordinary respondent cannot. For the sake of performing a rigorous test of the rationality hypothesis, we make the generous assumption that respondents know this correlation and use it.

If people make other-judgments by weighing their self-judgments with the overall (and factual) similarity between individual self-judgments and the group averages of the self-judgments over traits, their judgments of the average person will be regressive with respect to their self-judgments. Moreover, they will be regressive with respect to the group means. The result is a pattern of self-enhancement. To illustrate the emergence of self-enhancement from rational inference, we ran computer simulations with 10,000 cases each. Fig. 1 shows the results.

Each simulation comprised independent judgments with a preset variability for five traits, the mean judgments of which ranged from 2 to 6 in steps of 1. In the top left and center panels the dashed lines show these means. The solid lines show the means of the predictions. The error bars show the standard deviations, which were set to be .5 in the first simulation (top left) and 1.0 in the second (top center). The result of this increase in within-trait variability is a reduction in the correlation between individual judgments and means over traits, which in turn makes the predictions more regressive. Notice that in the second simulation (top center), the predicted line is shallower than in the first (top left). Assuming that positive traits receive higher self-judgments than do negative traits, the graphical representation of self-enhancement is the space between the actual and predicted line. For negative traits, judgments of others (predictions) tend to be higher than self-judgments (actual values). This is seen in the left half of each panel. For positive traits, the reverse is true. Judgments of others tend to be lower than self-judgments. The top right panel shows that increasing within-trait variability has little effect on the variability of the predictions.

The bottom panels show the effects of increasing within-trait variability on the correlation between the mean self-judgments and the predictions (left panel) as well as the intercept (center) and the regression slope (right). As the variability of the input increases, the predictions become more regressive, which results in greater self-enhancement.

The IRM predicts self-enhancement without appealing to self-enhancing motives or comparative processes (Krueger, 2000a; Moore & Healy, 2008; Moore & Swift, 2010), although it does not claim that such motives can never play a role. In the model, the assumed self-other similarity is an index of social projection. Hence, the model makes predictions of how changes in the strength of one "bias" affect changes in another. If social projection were less than the rational benchmark given by the correlation between self-judgments and their respective group means, judgments of others would be more regressive and self-enhancement would be stronger. Conversely, if projection were at its maximum ($r = 1$), self-enhancement would be eliminated but not reversed into self-effacement.[6]

One objective of the IRM is to provide testable point-hypotheses. In this example, a precise prediction was obtained with the multiplication rule, which shows what correlation to expect if both variables are correlated with a third variable (Alwin & Hauser, 1975). If, for example, the self-positivity correlation and the projection correlation are both .8, the expected correlation between group judgments and trait desirability judgments is .64, which yields a self-enhancement score of .16. Any empirical value greater than the expected IRM value can be said to represent "truly biased self-enhancement."[7]

3.2. Ingroup favoritism

To account for ingroup favoritism, the IRM begins with the assumptions that most people have positive self-images and that they project their own attributes more strongly onto ingroups than onto outgroups (Robbins & Krueger, 2005). The model predicts the positivity of the ingroup image as the product of self-positivity (the correlation between self-judgments and trait desirability judgments) and social projection to the ingroup (the correlation between self-judgments and group judgments). Likewise, it predicts the positivity of the outgroup image as the product of self-positivity and projection to the outgroup. The index of ingroup favoritism is the difference between the two products. Extending the numerical example introduced above with the assumption that projection to the outgroup is $r = .3$, the predicted magnitude of ingroup favoritism is .49. Again, although a difference is computed, it is not reified. There is no assumption of an ingroup favoritistic process, motive, or demon in the head.

3.3. Relations between phenomena

A crucial advantage of the IRM is that it permits the study of how discrete psychological phenomena are related

[6] For self-effacement to emerge, the self-image must be negative. Then, less projection yields more self-effacement.

[7] Could one argue that social projection emerges because of decreases in self-enhancement? Although this directional effect is possible, it is important to note that social projection is the simpler phenomenon; it is independent of trait desirability, it does not require processes of comparison (only of estimation), and specific processes of inductive reasoning have been shown to account for it (as opposed to a "need to self-enhance"). A strong theory favors whichever account requires fewer assumptions.

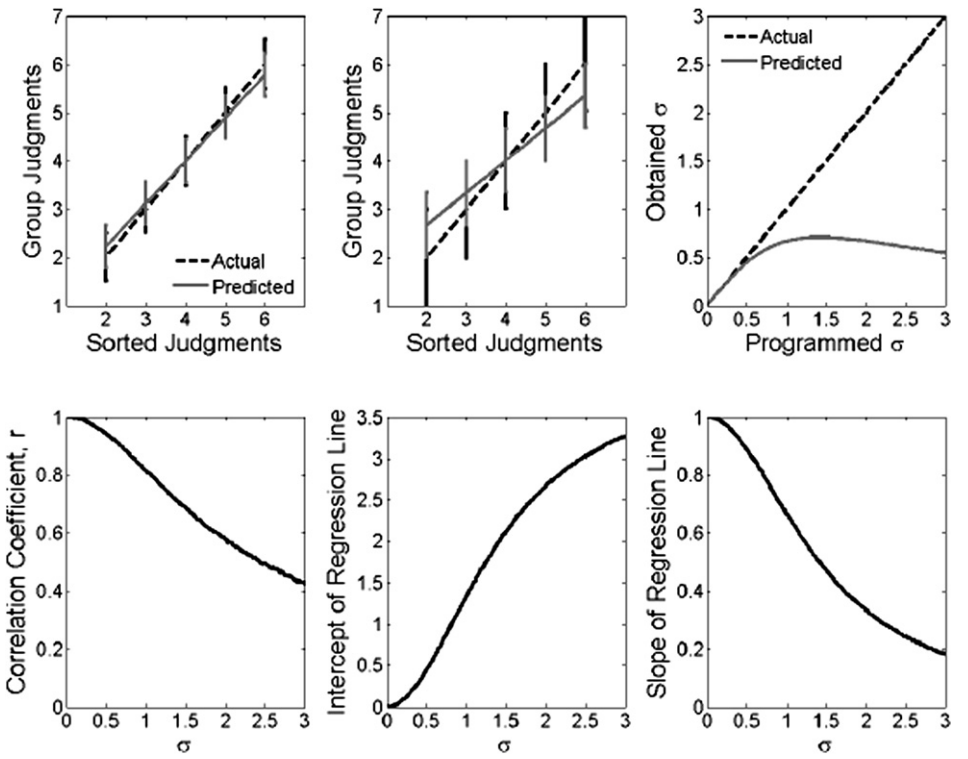

Fig. 1. Self-enhancement as a product of regression.

to one another. Such theoretical integration is a general *desideratum* (cf. Gheorghiu, Molz, & Pohl, 2004; Krueger, 2002). Since the IRM has derived a suite of point predictions from a small set of basic assumptions, it can predict and specify how the surface phenomena will be related to one another as they respond to changes in the input. These predicted relationships among the phenomena of interest can be empirically examined to determine whether the IRM derivations are sufficient to describe the data or if additional assumptions are needed. Finally, the model generates predictions that can be competitively tested against the predictions of other theories.

To illustrate the integrative potential of the model, consider the relationship between self-enhancement and ingroup favoritism. One influential theory suggests that self-enhancement and ingroup favoritism spring from the same motivational source: there, they should be positively correlated (Abrams & Hogg, 1988). The IRM shows that self-enhancement and ingroup favoritism can indeed be positively related and it shows how this happens. The critical variable is self-positivity. As self-images improve, *ceteris paribus*, both self-enhancement and ingroup favoritism increase. People who evaluate themselves very favorably are also those most likely to evaluate themselves more favorably than they evaluate others. Likewise, they are most likely to evaluate ingroups more favorably than outgroups, because they project more strongly to the former.

The IRM also shows, however, that if the degree of projection to the ingroup is varied while self-positivity is held constant, the correlation between self-enhancement and ingroup favoritism will be negative. As projection to the ingroup gets stronger, self-enhancement decreases because judgments of ingroups are less regressive (i.e., more highly correlated) with respect to self-judgments. At the same time, ingroup favoritism increases because judgments of outgroups become more regressive with respect to judgments of ingroups. This result is problematic for social identity theory.

Other theorists have suggested that self-enhancement and ingroup favoritism are independent (e.g., Brewer, 2007). Again, the IRM shows that such independence can occur and it identifies the conditions that make it so. Specifically, independence is expected when the effects of individual differences in self-positivity and the effects of individual differences in projection to the ingroup offset one another. This compensatory outcome will occur when increases in self-positivity are associated with decreases in projection to the ingroup. Although this pattern is theoretically possible, it is empirically rare.

For a theory that makes quantifiable predictions, computer simulations are a useful way to explore and display the range of possible outcomes before empirical study. Illustrating this feature for the IRM, the top panel of Fig. 2 shows the opposing effects of projection to the ingroup on self-enhancement and ingroup favoritism.

As projection to the ingroup increases (with projection to the outgroup being zero and with self-positivity being .5; all $\sigma = .2$), self-enhancement decreases whereas ingroup favoritism increases. The bottom panel shows the result of zooming in on specific levels of projection to the ingroup. At each level, the correlation between self-enhancement and ingroup favoritism is negative, but it becomes less so

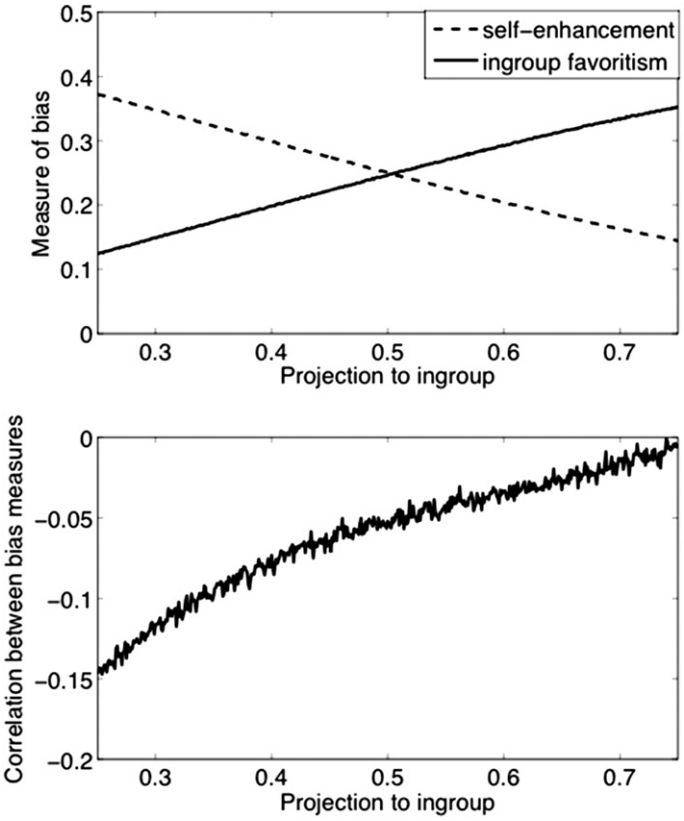

Fig. 2. The relationship between self-enhancement and ingroup favoritism.

as social projection increases. This is an example of a result that is not foreseen by any particular theory. Only the IRM could stimulate this simulation, and hence generate a novel prediction that is testable with empirical data.

Theory-based modeling is persuasive when the results provide a better fit to the data than modeling performed from an alternative theoretical perspective (Roberts & Pashler, 2000). In contrast to the IRM, self-categorization theory treats self-positivity as a secondary variable that can be predicted from the product of the ingroup's desirability and perceived group-self similarity (social projection or "self-stereotyping"). As the data do not show that self-positivity depends on perceived similarity (DiDonato et al., 2011), the IRM meets the demand of pitting some of its predictions against the predictions of competing theories, and winning.

The IRM does not reify the differences among correlations that represent the phenomena of self-enhancement and ingroup-favoritism (among others). Both phenomena emerge from a pattern of simple perceptions. Neither phenomenon is seen as a unitary belief arising from its own special motive, and neither requires an active comparative process. According to the conventional experimental perspective, ingroup-favoring individuals are rational because they engage in comparative reasoning. According to social-identity theory, these same individuals are irrational, also because they compare before they judge.

From the perspective of the IRM, no overall judgment of (ir)rationality is made. Still, the IRM permits a new look at correspondence rationality. Under most conditions, social projection increases the accuracy of judgments about groups (Hoch, 1987; Krueger, 1998). This is so because projection taps into the basic similarities among individuals. In the intergroup context, the model predicts that when differential projection (the difference between projection to the ingroup and projection to the outgroup) is greater than differential validity (the difference between the correlation between individual self-judgments and average self-judgments in the ingroup, and the correlation between individual self-judgments and average self-judgments in the outgroup), then there is evidence for an accuracy bias.

4. Toward stronger theories

A reviewer reminded us of Wittgenstein's (1958) observation that "in psychology there are experimental methods and *conceptual confusion*" (p. 232e; *italics* in the original). Wittgenstein goes on to write that "The existence of experimental methods makes us think we have the means of solving the problems which trouble us; though problem and methods pass each other by" (Wittgenstein, 1958). Attempting to respond to the great pessimistic philosopher, we note that strong theories look to the future, and doing so they satisfy a key criterion of rationality (Dawes, 1988; Krueger, 2000b). Strong theories make predictions over different samples, procedures, and measures. Strong theories capture the essence of induction, which is to anticipate

that which has not been experienced. When research questions become subtle, complex, or counterintuitive, forward-looking theories derive specific predictions mathematically or from computer simulations (Hastie & Stasser, 2000; McGuire, 1973). Ultimately, a convergence of theory, simulation, and empirical data is the strongest evidence that lessons have been learned. At the limit, the three levels of activity become indistinguishable. This idea can be traced to Turing (1950), and it has been used for model development in psychology (Church, 2001).

Strong theories account for existing data without overfitting them, which is what enables them to predict future data in a world in which the future is rarely exactly like the past (Pitt, Kim, & Myung, 2003). Strong theories are robust; they strike a balance between including enough parameters to achieve predictive success, while simultaneously limiting their number to a tolerable, parsimonious minimum (Gigerenzer, 2008; Hoffrage & Hertwig, 2012).

A call for stronger theories is timely in an area of psychological research that is still being viewed as soft, while at the same time it is an area that seeks to establish hard criteria for how ordinary people should think. Our first concern is the field's narrow focus on differences. Without a theory that predicts and explains when and how certain differences will emerge, empirical effects are likely to be redescribed and reified. In statistical terms, difference scores cannot explain anything that is not reducible to the effects of their components (McNemar, 1969). By definition, differences emerge as relationships. Reification is a fallacy; it forces empirical differences to explain themselves (Campbell, 1969).

In an earlier review of the judgment literature, Gigerenzer (1991) came to a similar conclusion, suggesting that a small set of heuristics that were proposed in the 1970s are little more than one-word re-descriptions of surface phenomena, that there are hardly any testable assumptions about how they work, and that because of their resulting over-elasticity, they can be used to explain phenomena that contradict each other (see Kahneman, 2003, for a more sanguine view).

To illustrate these problems with examples from well-known research programs, we focused on the simple difference score measures of self-enhancement and ingroup favoritism. More complex second-order difference scores are also popular. These scores are problematic for the same reasons we have discussed with reference to simple difference score measures, only more so. We alluded to Kwan's dual-difference score of self-enhancement as an example. The troubles of reified difference-score measures gain urgency when they entail morally charged inferences about (the lack of) rationality (De Martino, Kumaran, Seymour, & Dolan, 2006) or decency (Greenwald, McGhee, & Schwartz, 1998; see Blanton & Jaccard, 2006; Fiedler, Messner, & Bluemke, 2006; Ullrich, 2009, for psychometric critiques).

In developing strong theories, it is important to remember that a theory must not only describe the data it was developed to fit, it must also predict dependent measures and how these measures will respond to changes in the initial conditions. Theories that do this promise to uncover underlying processes, rather than simply provide a mathematical description of the phenomenon. To develop such theories is a challenge for psychologists in all fields, and it is rarely addressed in the literature (see Guilhardi & Church, 2009, for a pertinent example).

To overcome the limitations of shallow, effect-driven theorizing, we introduced the IRM as a way to predict and explain several empirical phenomena. A central psychological process assumed by the IRM is social projection. We asked whether social projection might itself be a reifying label attached to an empirical finding, and we concluded that this position is difficult to hold. Decades of research have shown that social projection can be understood as a special case of inductive reasoning (Dawes, 1989; Krueger, 2000a).

A critique of our approach may take the form that the IRM commits the sins it is designed to overcome. As the model relies on statistical regression and its attendant effects, it must show that it can explain phenomena such as self-enhancement or ingroup favoritism *without explaining them away*. Should these effects perhaps be regarded as artifacts or epiphenomena? Again, we submit that the answer is no. Findings are artifactual if they emerge from a lack of internal validity. That is, findings are artifactual if they are produced by method alone, not by the forces the methods are meant to represent. Phenomenologically, self-enhancement is quite real, and the IRM offers a sufficient explanation based on the interplay of two basic phenomena: a positive self-image and projective judgments about others. The model makes clear that the latter must be regressive in order to be rational. What is more, the IRM can predict variations in the strength of self-enhancement. Hence, there is no artifact.

If self-enhancement and ingroup favoritism are not artifacts, might they still be dismissed as epiphenomena? An epiphenomenon is caused, like everything else, by a specific pattern of antecedent conditions, but it has no effects of its own. An epiphenomenon is a dead end. Wegner (2002), for example, claimed that the experience of conscious will is illusory and epiphenomenal. Behavior would continue to flow as it does now if there were no consciousness. In contrast, ingroup favoritism predicts discriminatory behavior in favor of the ingroup. In other words, ingroup favoritism is part of a causal chain running from self-image and projection to discrimination. Removing the link of ingroup favoritism would break the chain.

Our second concern is that an experimental focus on differences can hamper efforts to think clearly about human rationality. To repeat: In some cases, differences resulting from comparative reasoning are taken as evidence of irrational reasoning. In other cases, multiple differences are proposed and compete for the distinction of being the true measure. In still other cases, surface differences obscure several different underlying differences, thus complicating interpretation. A theory that derives differential responses from basic psychological premises takes the heat off the rationality issue. Outcome measures such as self-enhancement or ingroup favoritism can be regarded dispassionately. If there is no self-enhancing or ingroup-enhancing motive or process, these outcomes by themselves are perhaps best regarded *arational*.

The question of rationality may be more poignantly asked with respect to the underlying processes. In the case of the IRM, social projection is the critical underlying process, playing a central role in generating the predictions the model makes. As noted above, projection is a useful heuristic that works through association. Within limits, it allows humans to predict that which they have not experienced. Good scientific theories do the same.

References

Abrams, D., & Hogg, M. A. (1988). Comments on the motivational significance of self-esteem in social identity and intergroup discrimination. *European Journal of Social Psychology, 18*, 317–334. http://dx.doi.org/10.1002/ejsp.2420180403.

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

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40*, 37–47. http://dx.doi.org/10.2307/2094445.

Bakker, M., & Wicherts, J. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*1–13. http://dx.doi.org/10.1002/ejsp.2420180403, online first.

Bargh, J. A. (2007). Social psychological approaches to consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 555–569). New York, NY: Cambridge University Press.

Bartsch, R. A., & Judd, C. M. (1993). Majority – minority status and perceived ingroup variability revisited. *European Journal of Social Psychology, 23*, 471–483. http://dx.doi.org/10.1002/ejsp.2420230505.

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology,* . http://dx.doi.org/10.1037/a0021524.

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist, 61*, 27–41. http://dx.doi.org/10.1037/0003-066X.61.1.27.

Brewer, M. B. (1991). The social self: on being the same and different at the same time. *Personality and Social Psychology Bulletin, 17*, 475–482. http://dx.doi.org/10.1177/0146167291175001.

Brewer, M. B. (2007). The importance of being *we*: human nature and intergroup relations. *American Psychologist, 62*, 728–738. http://dx.doi.org/10.1037/0003-066X.62.8.728.

Brown, J. D. (2012). Understanding the better than average effect: motives (still) matter. *Personality and Social Psychology Bulletin, 38*, 209–219. http://dx.doi.org/10.1177/0146167211432763.

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*, 171–189. http://dx.doi.org/10.1006/jmps.1999.1282.

Campbell, D. T. (1969). Perspective: artifact and control. In R. Rosenthal, & R. Rosnow (Eds.), *Artifact in behavior research* (pp. 351–382). New York, NY: Academic Press.

Chambers, J. R., & Windschitl, P. P. D. (2004). Biases in comparative judgments: the role of nonmotivated factors in above-average and comparative optimism effects. *Psychological Bulletin, 130*, 813–838. http://dx.doi.org/10.1037/0033-2909.130.5.813.

Church, R. M. (2001). A Turing test for computational and associative models of learning. *Current Directions in Psychological Science, 10*, 132–136. http://dx.doi.org/10.1111/1467-8721.00133.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: negative implications for mental health. *Journal of Personality and Social Psychology, 68*, 1152–1162. http://dx.doi.org/10.1037/0022-3514.68.6.1152.

Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research, 17*, 143–169. http://dx.doi.org/10.1023/B:SORE.0000027407.70241.15.

Dawes, R. M. (1988). *Rational choice in an uncertain world*. San Diego, CA: Harcourt, Brace, Jovanovich.

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology, 25*, 1–17. http://dx.doi.org/10.1016/0022-1031(89)90036-X.

Dawes, R. M. (2001). *Everyday irrationality: How pseudo-scientists, lunatics, and the rest of us systematically fail to think rationally*. Boulder, CO: Westview Press.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313*, 684–687. http://dx.doi.org/10.1126/science.1128356.

DiDonato, T. E., Ullrich, J., & Krueger, J. I. (2011). Social perception as induction and inference: an integrative model of intergroup differentiation, ingroup favoritism, and differential accuracy. *Journal of Personality and Social Psychology, 100*, 66–83. http://dx.doi.org/10.1037/a0021051.

Edwards, J. R., & Berry, J. W. (2010). The presence of something or the absence of nothing: increasing theoretical precision in management research. *Organizational Research Methods, 13*, 668–689. http://dx.doi.org/10.1177/1094428110380467.

Ellsworth, P., & Gonzalez, R. (2003). Questions and comparisons: methods of research in social psychology. In M. A. Hogg, & J. Cooper (Eds.), *The Sage handbook of social psychology* (pp. 24–42). Thousand Oaks, CA: Sage.

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS One, 5*(4), e10068. http://dx.doi.org/10.1371/journal.pone.0010068.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics, 90*, 891–904. http://dx.doi.org/10.1007/s11192-011-0494-7, Published online: 11 September.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140. http://dx.doi.org/10.1177/001872675400700202.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*, 193–214. http://dx.doi.org/10.1037/0033-295X.103.1.193.

Fiedler, K. (2011). Voodoo correlations are everywhere—not only in neuroscience. *Perspectives on Psychological Science, 6*, 163–171. http://dx.doi.org/10.1177/1745691611400237.

Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α-control to validity proper: problems with a short-sighted false-positive debate. *Perspectives on Psychological Science, 7*, 661–669. http://dx.doi.org/10.1177/1745691612462587.

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T": a logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology, 17*, 74–147. http://dx.doi.org/10.1080/10463280600681248.

Fraley, R. C., & Marks, M. J. (2007). The null hypothesis significance testing debate and its implications for personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 149–169). New York: Guilford.

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General, 141*, 124–133. http://dx.doi.org/10.1037/a0024006.

Gallistel, C. R. (2009). On the importance of proving the null. *Psychological Review, 116*, 439–453. http://dx.doi.org/10.1037/a0015251.

Gheorghiu, V. A., Molz, G., & Pohl, R. (2004). Suggestion and illusion. In R. Pohl (Ed.), *Cognitive illusions: A handbook of biases in thinking, judgement and memory* (pp. 399–421). New York, NY: Psychology Press.

Gigerenzer, G. (1991). From tools to theories: a heuristic of discovery in cognitive psychology. *Psychological Review, 98*, 254–267. http://dx.doi.org/10.1037/0033-295X.98.2.254.

Gigerenzer, G. (2008). *Rationality for mortals*. New York, NY: Oxford University Press.

Gilovich, T. (1997). Some systematic biases of everyday judgment. *Skeptical Inquirer, 21*, 31–35.

Gilovich, T., & Griffin, D. W. (2010). Judgment and decision making. In S. T. Fiske, D. T. Gilbert, & G. Lindsey (Eds.), *Handbook of social psychology* (5th ed., Vol. 1, pp. 542–588). Hoboken, NJ: Wiley.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480. http://dx.doi.org/10.1037/0022-3514.74.6.1464.

Grice, H. P. (1975). Logic of conversation. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts*. New York, NY: Wiley.

Guenther, C. L., & Alicke, M. D. (2010). Deconstructing the better-than-average effect. *Journal of Personality and Social Psychology, 99*, 755–770. http://dx.doi.org/10.1037/a0020959.

Guilhardi, P., & Church, R. M. (2009). The generality of empirical and theoretical explanations of behavior. *Behavioural Processes, 81*, 205–215. http://dx.doi.org/10.1016/j.beproc.2009.01.013.

Hastie, R., & Dawes, R. M. (2010). *Rational choice in an uncertain world: The psychology of judgment and decision making* (2nd ed.). Thousand Oaks, CA: Sage.

Hastie, R., & Stasser, G. (2000). Computer simulation methods for social psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 85–115). New York, NY: Cambridge University Press.

Herbert, W. (2010). *On second thought: Outsmarting your mind's hard-wired habits*. New York, NY: Random House.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind. Improving individual judgments with dialectical bootstrapping.

Psychological Science, 20, 231–237. http://dx.doi.org/10.1111/j.1467-9280.2009.02271.x.

Higgins, E. T. (2004). The eighth koan of progress in social psychology: a variable anointed as "special" will demand special treatment. In J. T. Jost, M. R. Banaji, & D. A. Prentice (Eds.), *Perspectivism in social psychology: The yin and yang of scientific progress* (pp. 305–317). Washington, DC: American Psychological Association.

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: the pros and cons of social projection. *Journal of Personality and Social Psychology, 53*, 221–234. http://dx.doi.org/10.1037/0022-3514.53.2.221.

Hoffrage, U., & Hertwig, R. (2012). Simple heuristics in a complex social world. In J. I. Krueger (Ed.), *Social judgment and decision making* (pp. 135–150). New York, NY: Psychology Press.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. New York, NY: Basic Books.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine, 2*, 696–701. http://dx.doi.org/10.1371/journal.pmed.0020124.

Jacoby, L. L. (1991). A process dissociation framework: separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541. http://dx.doi.org/10.1016/0749-596X(91)90025-F.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206–219. http://dx.doi.org/10.1037/0022-3514.66.1.206.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American Psychologist, 58*, 697–720. http://dx.doi.org/10.1037/0003-066X.58.9.697.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11*, 123–141. http://dx.doi.org/10.1016/0010-0277(82)90022-1.

Karniol, R. (2003). Egocentrism versus protocentrism: the status of self in social prediction. *Psychological Review, 110*, 564–580. http://dx.doi.org/10.1037/0033-295X.110.3.564.

Keeley, B. (2007). God as the ultimate conspiracy theory. *Episteme, 4*, 135–149. http://dx.doi.org/10.1353/epi.2007.0021.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.). *Nebraska symposium on motivation, Vol. 15* (pp. 192–238). Lincoln, NE: University of Nebraska Press.

Kenrick, D. T. (1994). Evolutionary social psychology: form sexual selection to social cognition. *Advances in Experimental Social Psychology, 26*, 75–121.

Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review, 2*, 196–217. http://dx.doi.org/10.1207/s15327957pspr0203_4.

Keynes, J. M. (1936). *The general theory of employment, interest and money*. New York, NY: Harcourt Brace and Co.

Klar, Y., & Giladi, E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin, 25*, 586–595. http://dx.doi.org/10.1177/0146167299025005004.

Krueger, J. (1998). On the perception of social consensus. In M. P. Zanna (Ed.). *Advances in experimental social psychology, Vol. 30* (pp. 163–240). San Diego, CA: Academic Press.

Krueger, J. (2000a). The projective perception of the social world: a building block of social comparison processes. In J. Suls, & L. Wheeler (Eds.), *Handbook of social comparison: Theory and research* (pp. 323–351). New York: Plenum/Kluwer.

Krueger, J. (2000b). Distributive judgments under uncertainty: Paccioli's game revisited. *Journal of Experimental Psychology: General, 129*, 546–558. http://dx.doi.org/10.1037/0096-3445.129.4.546.

Krueger, J. (2001). Null hypothesis significance testing: on the survival of a flawed method. *American Psychologist, 56*, 16–26. http://dx.doi.org/10.1037/0003-066X.56.1.16.

Krueger, J. (2002). On the reduction of self-other asymmetries: benefits, pitfalls, and other correlates of social projection. *Psychologica Belgica, 42*, 23–41.

Krueger, J. I. (2003). Return of the ego—self-referent information as a filter for social prediction: comment on Karniol (2003). *Psychological Review, 110*, 585–590. http://dx.doi.org/10.1037/0033-295X.110.3.585.

Krueger, J. I. (2008). The robust beauty of simple associations. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn M. Dawes* (pp. 111–140). New York, NY: Psychology Press.

Krueger, J. I. (2009). A componential model of situation effects, person effects and situation-by-person interaction effects on social behavior. *Journal of Research in Personality, 43*, 127–136. http://dx.doi.org/10.1016/j.jrp.2008.12.042.

Krueger, J. I., Acevedo, M., & Robbins, J. M. (2006). Self as sample. In K. Fiedler, & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 353–377). New York, NY: Cambridge University Press.

Krueger, J. I., & DiDonato, T. E. (2008). Social categorization and the perception of groups and group differences. *Social and Personality Psychology Compass: Group Processes, 2*, 733–750. http://dx.doi.org/10.1111/j.1751-9004.2008.00083.x.

Krueger, J. I., Freestone, D., & Heck, P. *The inductive reasoning model: Rationale and simulation*. Brown University, working paper, in preparation.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27*, 313–376. http://dx.doi.org/10.1017/S0140525X04000081.

Krueger, J. I., & Wright, J. C. (2011). Measurement of self-enhancement (and self-protection). In M. D. Alicke, & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 472–494). New York, NY: Guilford.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review, 118*, 97–109. http://dx.doi.org/10.1037/a0023709.

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: an interpersonal approach. *Psychological Review, 111*, 94–110. http://dx.doi.org/10.1037/0033-295X.111.1.94.

Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge, UK: Cambridge University Press.

Latrofa, M., Vaes, J., Cadinu, M., & Carnaghi, A. (2010). The cognitive representation of self-stereotyping. *Personality and Social Psychology Bulletin, 36*, 911–922. http://dx.doi.org/10.1177/0146167210373907.

Locke, J. (1690/1970). *An essay concerning human understanding*. Menston: Scolar Press.

McCauley, C. R., & Stitt, C. L. (1978). An individual and quantitative measure of stereotypes. *Journal of Personality and Social Psychology, 36*, 929–940. http://dx.doi.org/10.1037/0022-3514.36.9.929.

McGuire, W. J. (1973). The yin and yang of progress in social psychology. *Journal of Personality and Social Psychology, 26*, 446–456. http://dx.doi.org/10.1037/h0034345.

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York, NY: Wiley.

Meehl, P. E. (1956). Wanted—a good cookbook. *American Psychologist, 11*, 263–272. http://dx.doi.org/10.1037/h0044164.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*, 806–834. http://dx.doi.org/10.1037/0022-006X.46.4.806.

Meehl, P. E. (1990). Appraising and amending theories: the strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry, 1*, 108–141. http://dx.doi.org/10.1207/s15327965pli0102_1.

Mill, J. S. (1872/1973). System of logic. In J. M. Robson (Ed.), *Collected works of John Stuart Mill*, (8th ed.) Vols. VII–VIII University of Toronto Press, (Original edition published in 1872).

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*, 102–117. http://dx.doi.org/10.1037/0033-295X.115.2.502.

Moore, D. A., & Swift, S. A. (2010). The three faces of overconfidence in organizations. In D. De Cremer, R. van Dick, & J. K. Murnighan (Eds.), *Social psychology and organizations* (pp. 147–184). New York, NY: Routledge.

Musil, R. (1930). *Der Mann ohne Eigenschaften, Bd. 1*. Berlin: Rowohlt.

Mussweiler, T. (2003). Comparison processes in social judgment: mechanisms and consequences. *Psychological Review, 110*, 472–489. http://dx.doi.org/10.1037/0033-295X.110.3.472.

Nagel, R. (1995). Unraveling in guessing games: an experimental study. *American Economic Review, 85*, 1313–1326.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Park, B., & Rothbart, M. (1982). Perception of out-group homogeneity and levels of social categorization: memory for the subordinate attributes of in-group and out-group members. *Journal of Personality and Social Psychology, 42*, 1051–1068. http://dx.doi.org/10.1037/0022-3514.42.6.1051.

Payne, B. K., & Iannuzzi, J. L. B. (2012). Automatic and controlled decision making: a process dissociation perspective. In J. I. Krueger (Ed.), *Social judgment and decision-making* (pp. 41–58). New York, NY: Psychology Press.

Pitt, M. A., Kim, W., & Myung, I. J. (2003). Flexibility versus generalizability in model selection. *Psychnomic Bulletin & Review, 10*, 29–44. http://dx.doi.org/10.3758/BF03196467.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London, UK: Routledge.

Quetelet, A. (1835/1968). *A treatise on man and the development of his faculties*. New York, NY: B. Franklin.

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: a review and meta-analysis. *Personality and Social Psychology Review, 9*, 32–47. http://dx.doi.org/10.1207/s15327957pspr0901_3.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*, 358–367. http://dx.doi.org/10.1037/0033-295X.107.2.358.

Runkel, P. J., & McGrath, J. E. (1974). *Research on human behavior: A systematic guide to method*. New York, NY: Holt, Rinehart, & Winston.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366. http://dx.doi.org/10.1177/0956797611417632.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *Psychology of intergroup relations* (pp. 33–47). Chicago: Nelson-Hall.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460. http://dx.doi.org/10.1093/mind/LIX.236.433.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. (1987). *Rediscovering the social group: A self-categorization theory*. Oxford, England: Basil Blackwell.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124–1130. http://dx.doi.org/10.1126/science.185.4157.1124.

Ullrich, J. (2009). Reconsidering the "relative" in relative ingroup prototypicality. *European Journal of Social Psychology, 39*, 299–310. http://dx.doi.org/10.1002/ejsp.540.

Van Lange, P. A. M., Higgins, E. T., & Kruglanski, A. (2011). *Handbook of theories of social psychology*. New York, NY: Sage.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12*, 129–140. http://dx.doi.org/10.1080/17470216008416717.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Windschitl, P., Rose, J. P., Stalkfleet, M. T., & Smith, A. R. (2008). Are people excessive or judicious in their egocentrism? A modeling approach to understanding bias and accuracy in people's optimism. *Journal of Personality and Social Psychology, 95*, 253–273. http://dx.doi.org/10.1037/0022-3514.95.2.253.

Wittgenstein, L. (1958). *Philosophical investigations* (translated by G.E.M. Ascombe) (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Wundt, W. M. (1874). *Grundzüge der physiologischen Psychology* [*Elements of physiological psychology*]. Leipzig, Germany: Engelmann.

Yzerbyt, V., & Demoulin, S. (2010). Intergroup relations. In S. T. Fiske, D. T. Gilbert, & G. Lindsey (Eds.), *Handbook of social psychology* (5th ed., Vol. 2, pp. 1024–1083). Hoboken, NJ: Wiley.