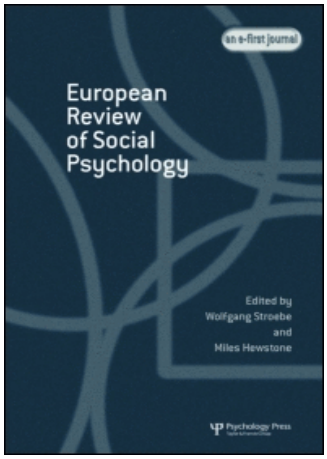

European Review of Social Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713684724

From social projection to social behaviour

First Published on: 01 January 2007
To cite this Article: Krueger, Joachim I. (2007) 'From social projection to social behaviour', European Review of Social Psychology, 18:1, 1 - 35
To link to this article: DOI: 10.1080/10463280701284645
URL: http://dx.doi.org/10.1080/10463280701284645

EUROPEAN REVIEW OF SOCIAL PSYCHOLOGY
2007, 18, 1–35

Psychology Press
Taylor & Francis Group

From social projection to social behaviour

Joachim I. Krueger
Brown University, Providence, RI, USA

Social projection is a judgemental heuristic that allows people to make quick and reasonably accurate predictions about others. The first part of this paper presents a review of the status of projection as a highly (though not fully) automatic process, its separateness from superficially similar processes of self-stereotyping, and its implications for intergroup perception. The second part places social projection within the context of the theory of evidential decision making, which highlights the benefits and the liabilities of projection in social dilemma situations. The main benefit is that projection can enhance cooperation within a group by leading individuals to believe that their own behavioural choices will be reciprocated. However, when interpersonal social dilemmas are nested within intergroup dilemmas, differential projection (i.e., strong ingroup projection paired with weak outgroup projection) yields collectively undesirable outcomes.

> From our own case we believe in that which we do not know.
>
> Augustine, *De Trinitate* 8.6.9

The concept of social projection is once again generating vigorous theory development and empirical research in social psychology. This attention is deserved because social projection is among the simplest, oldest, and arguably most central concepts of the field. It is simple: People by and large expect that others are similar to them. It is old: F. H. Allport (1924) anchored his analysis of crowd behaviour on the idea of projection. It is

Correspondence should be addressed to Joachim I. Krueger, Department of Psychology, Brown University, Box 1853, 89 Waterman Street, Providence, RI 02912, USA.
E-mail: Joachim_Krueger@Brown.edu

I am grateful to Maya Machunsky and Thorsten Meiser for their encouragement, support, and perceptive comments. Likewise, I am indebted to three anonymous reviewers who helped me improve this manuscript with their constructive suggestions. As usual, Judith Schrier was generous with her editorial feedback.

http://www.psypress.com/ersp DOI: 10.1080/10463280701284645

central: Without social projection, social intelligence and the effectiveness of social behaviour would be diminished.

Social projection may be defined as the process by which people come to believe that others are similar to them. This definition goes beyond the long-held view that projection can be defined in terms of its outcome, namely a positive correlation between judgements about the self and judgements about others. A definition that refers to mental processes must guide the empirical study of these processes. This endeavour has proven difficult. Once the basic correlational finding was established, investigators began to ask how it could be produced experimentally. A variety of mental processes, mostly those having to do with the selective or privileged processing of self-referent information, emerged as contributors to the correlation between self-judgements and other judgements. The ironic result was that once social projection was observed in a data set, it became less clear which of these processes was most responsible. The greater the number of available and sufficient causes, the lower is the probability that any particular cause is operative. A second result was that the various causes only increased the size of the projective correlation; when any one of these causes was absent, some projection still occurred. When reviewing this literature a decade ago, I concluded that social projection is a perceptual primitive that emerges with minimal cognitive contribution (Krueger, 1998). In this paper, I revisit this conclusion because some new evidence suggests that projection can be engaged and suspended strategically.

The paper is divided in two major parts. In part one, I review the current status of three issues. The first issue is the degree to which projection is automatic. The second issue is the conflict between social projection and self-stereotyping. The third issue is the moderating effect of social categorisation, and its implications for intergroup perception. As we shall see, a neat separation of these issues is not possible, and some cross-referencing will be necessary.

In part two, I suggest that social projection can be understood within the theory of evidential decision making. The goal of this theory is to model rational choice between alternatives in social dilemma situations, where self-judgements and other judgements are reciprocally determined. Again, I focus on three issues. The first issue is the logic of evidential decision making and its relation to classic game theory. I argue that the evidential theory can account for cooperation, whereas the classic theory cannot. The second issue is how the theory compares with competing theories stressing the role of social preferences. Again, I argue that the evidential theory explains cooperation most successfully. The third issue is the implications of social projection for intergroup behaviour and conflict when social projection is moderated by social categorisation. Here, I argue

that the social projection hypothesis is more parsimonious than theories postulating motivational differences separating individuals and groups.

CONTEMPORARY ISSUES IN THE SEARCH FOR PROJECTIVE PROCESSES

Automaticity and control

A prototypically automatic process occurs outside awareness, requires no effort or intention, and cannot be stopped at will (Bargh, 1994; Moors & de Houwer, 2007). This multi-facetedness of automaticity makes it unlikely that a critical experiment will yield a final verdict as to whether projection is automatic or controlled. However, a body of evidence suggests that projection is highly automatic most of the time.

Awareness does not appear to be necessary for projection. Many research participants deny that they generalise from themselves to others even when their own responses show that they do. Yet some of the same respondents feel that other individuals generalise their responses to the group. In other words, these participants have insight into the reality of projection, but fail to attribute it to themselves (Krueger & Zeiger, 1993). There is a certain irony in this meta-awareness because it amounts to an exception to the otherwise automatic appearance of projection. If people thought they themselves did not project, this belief should also be projected onto others.

Further evidence for automaticity comes from studies on nonconscious defensive projection. In a neo-Freudian vein, Newman, Duff, Kimberly, and Baumeister (1997) reasoned that people project their own personal attributes onto others when they are trying not to think about these attributes. When people try not to think about their own undesirable traits, they tend to succeed, at least for a while. The inhibition of awareness consumes mental resources, however, and it makes the suppressed material hyperaccessible. When others are being judged, any readily accessible material plays a prominent role (Govorun, Fuegen, & Payne, 2006; Schimel, Greenberg, & Martens, 2003). An alternative method of studying automatic projection is to present the to-be-projected information subliminally. Kawada, Oettingen, and Gollwitzer (2004) found that participants who were primed with the goal to compete were more likely than controls to perceive others as competitors.

The sufficiency of perceptual priming as a trigger of projection suggests that the process does not require effort. People do not have to think hard to reach the assumption that others are similar to them. Self-referent information is readily available for many issues that people confront. Of course,

there are stimulus effects. It may be harder to decide whether one favours French or Italian movies than it is to choose between skiing or skydiving as a Sunday afternoon activity. The question is whether projection is disrupted when judgements become more difficult. This does not appear to be the case. In one study, projection was just as strong when participants were under high cognitive load as when they were not (Krueger & Stanke, 2001), and in another, time pressure even increased projection (Epley, Keysar, & van Boven, 2004).

Intention and control are the two sides of the agency coin. The question of whether people only project when they intend to has received little attention. In part, the evidence for nonconscious and effortless projection implies that intentions are unnecessary. However, intention may be sufficient to increase projection. To my knowledge, this question has not been examined. If such tests were successful, they would presumably be dismissed as demonstrations of demand characteristics. Control, on the other hand, raises the question of whether people can deliberately abstain from projection. One debiasing study showed that simple forewarnings have no effect (Krueger & Clement, 1994), although recent evidence shows that incentives for accuracy reduce egocentrism (Epley et al., 2004). In communication, people fail to set aside information that only they, but not their interlocutors, have. Speakers often refer to such information as though their audience knew about it. As a consequence, they overestimate the effectiveness of their own communication (Keysar, Lin, & Barr, 2003).

On balance, it seems that social projection is a primitive and robust phenomenon—it can operate without awareness or effort, it does not require intention, and it does not respond well to attempts at curbing it. There is, however, a striking exception. When the social target is an outgroup, projection breaks down (Robbins & Krueger, 2005). There are competing accounts for how this happens. One possibility is that projection is not triggered in the first place when people recognise the fact that they are not included in the group. Alternatively, projection to the outgroup is automatically engaged, but then reduced by a deliberate and time-consuming adjustment process (Epley et al., 2004; Krueger, 2000).

The anchoring-and-adjustment hypothesis was supported in a study with minimal groups (DiDonato & Krueger, 2007). Participants were simultaneously categorised as lovers of Klee or Kandinsky paintings and as over- or under-estimators of dots. Participant then judged themselves and members of three groups with regard to a variety of attitude statements. As expected, they projected strongly to members of their double ingroup and barely to members of the double outgroup. The moderate level of projection to the mixed group was consistent with a sliding process of adjustment. Projection coefficients for the mixed group were no more variable than coefficients for the pure ingroup or outgroup. The intermediate

effect size was thus not an artifact of some participants projecting fully to the mixed group and others not projecting at all.[1]

Ames (2004a, 2004b) proposed that projection is strategically regulated. According to his similarity-contingency model, people deploy or withhold projection depending on the outcome of a preliminary assessment of similarity. When a person or group appears to be similar to the self, people generate further expectations of similarity by projecting other attributes of their self-concept. In contrast, when a person or group appears to be dissimilar, people generate expectations by using social stereotypes if such stereotypes are available. The similarity-contingency model is appealing in that it provides a common platform for projection and stereotyping. Both represent, after all, some form of inductive reasoning that fills in missing social information. By allowing self-regulated social perception, the model also overcomes the limitations of a pure automaticity hypothesis. Nonetheless, the model fails to explain how the initial similarity judgements arise if not from social categorisation. In as much as the classification of people into broad categories such as gender, age, race, or ethnicity is either biologically prepared or socially overlearned, it provides the basis for both global similarity judgements and attribute-by-attribute projection. In the laboratory, perceptions of high or low similarity can be induced, but they require the provision of specific person information. In other words, such perceptions require the very kind of information that is supposed to be the domain of projection and stereotyping.

Social projection vs self-stereotyping

Positive correlations between self-judgements and group judgements are necessary for demonstrations of social projection. They are not sufficient, however. Arguably, such correlations can reflect the reverse causal path. Under certain conditions, people may select their own responses depending on what they believe to be the response of the majority. Social behaviour is known to be open to conformity effects. People often hop on a bandwagon when they find it gratifying to be in the majority or because they perceive the majority behaviour as a valid cue towards the best or most accurate response (Cialdini & Trost, 1998). Likewise, self-categorisation theory suggests that self-perception can be shaped by what a person believes to be the majority attribute in the ingroup (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). Under certain conditions, people are thought to perceive themselves in terms of their group membership rather than in terms of unique personal attributes. Thus, perceived similarities between

[1]See Crisp and Hewstone (2007) for a review of research on crossed categorisation.

self-judgements and group judgements are thought to reflect processes of depersonalisation and self-stereotyping.

Additional empirical data are needed to clarify the directional path of observed correlations between self and group judgements. My colleagues and I recently conducted a comparative literature review, and found a preponderance of evidence for social projection (Krueger, Acevedo, & Robbins, 2005). The following list is an overview of seven pieces of evidence favouring the social projection hypothesis.

1. *Response time*: Self-judgements are faster than group judgements (Clement & Krueger, 2000).
2. *Response facilitation*: Self-judgements facilitate subsequent group judgements more than vice versa (Clement & Krueger, 2000).
3. *Variability*: Self-judgements induce greater variability in subsequent group judgements than vice versa (Dawes, McTavish, & Shacklee, 1977).
4. *Malleability*: Self-judgements are more stable than are group judgements, and are more resistant to experimentally induced change (Krueger & Stanke, 2001).
5. *Self-reports*: People find self-judgements to be easier than group judgements, and they make them with greater confidence (Krueger & Stanke, 2001).
6. *Disambiguity*: Social projection is observed when no stereotypes exist (i.e., in minimal laboratory groups), whereas no comparable case for unambiguous self-stereotyping exists (Clement & Krueger, 2002).
7. *Purity of social categorisation*: Correlations between self-judgements and ingroup judgements are larger in minimal groups than in real groups, whereas self-stereotyping would require the opposite (Robbins & Krueger, 2005).

An eighth piece of evidence may be added, namely the general finding that stereotypes affect judgements of an individual only to the degree that little else is known him or his/her (Krueger & Rothbart, 1988; Kunda & Sherman-Williams, 1993). Stereotyping oneself ought to be harder than stereotyping others because it requires the displacement of a larger amount of person-specific information. For the same reason, it is easier to stereotype members of outgroups than members of ingroups (Ames, 2004b).

However compelling the circumstantial evidence may be, it cannot replace a critical experiment. Using a modified minimal group paradigm, one study directly compared the strength of self-stereotyping with the strength of social projection. Cadinu and Rothbart (1996) provided participants either with selective information about a group to which they belonged, or with information about themselves. They then assessed

respectively how much participants generalised from the ingroup to themselves and how much they generalised from themselves to the ingroup. Both effects were statistically significant, but the latter was about twice as large as the former.

Using a response-time measures, Otten and Epstude (2006) found that dichotomous judgements (e.g., "describes me [my group]" vs "does not describe me [my group]") were faster when the responses for the self and the group were the same than when they were different. The critical evidence for social projection was that this pattern emerged even for traits on which numerical group judgements—but not self-judgements—indicated indifference (i.e., a rating of 4 on a 7-point scale). When participants were forced to decide whether a trait applied to their group, they used their self-judgements as anchors. When this method was reversed to examine self-stereotyping, no such effect emerged. Dichotomous self-judgements for traits that participants had placed at the midpoint of the scale were not assimilated to the pre-existing, non-neutral group judgements.

A possible objection to these comparisons between social projection and self-stereotyping is that they gloss over the contextual constraints on the latter. If self-stereotyping occurs only under specific conditions, its overall effect size will underestimate the true effect in the conditions that matter. Self-categorisation theory is not entirely clear about what these conditions are. Our literature review suggested the following: Social categorisation has to be salient, the person has to be highly identified with the ingroup, the individual self has to be threatened, and the attributes in question have to be evaluatively charged (by most accounts they have to be positive; Krueger et al., 2005). At present, there are not enough studies that have simultaneously varied several of these conditions. However, the review suggests that none of these variables is singly sufficient to elicit self-stereotyping strong enough to override projection. Even studies designed to elicit self-stereotyping show that concurrent projection effects are stronger (e.g., Brewer & Weber, 1994).

The ease with which projection is empirically produced is noteworthy in light of a statistical asymmetry suggesting that is should be otherwise. Dawes (1990) showed that inferences from data to the category from which they were sampled have a smaller error variance than inferences from categories to data. This asymmetry suggests that social projection should harder to demonstrate than self-stereotyping.

Social projection and social categorisation

As noted earlier, social categorisation moderates social projection. Self-judgements predict judgements about ingroups much better ($r \approx .5$) than they

predict judgements about outgroups ($r \approx .1$; Robbins & Krueger, 2005). This difference has important implications for intergroup perception. Consider four standard findings. First, people tend to accentuate (i.e., exaggerate) differences between groups (Krueger, 1992). Second, people tend to perceive outgroups as more homogeneous than ingroups (and show the reverse tendency under certain conditions; Rubin & Badea, 2007). Third, people favour ingroups over outgroups both perceptually and behaviourally (Brewer, 1999). Fourth, perceptions of outgroups tend to be less accurate than perceptions of ingroups (Ryan & Bogart, 2001).

Research on these phenomena has been stimulated by the theoretical ideas and empirical results presented over several decades by Henri Tajfel and his colleagues. Whereas some researchers favour motivational explanations, others consider cognitive accounts to be more parsimonious. Tajfel's own perspective changed several times. Inspired by the New Look on Perception, his early work emphasised the effects of psychological needs on the perception of objects (Tajfel, 1959). Later, he sought to explain the perceptual exaggeration of intergroup differences and the minimisation of intragroup differences within a more cognitive framework (Tajfel, 1969). With social-identity theory, he eventually brought back assumptions about hedonic needs to explain ingroup bias and intergroup discrimination (Tajfel & Turner, 1979).

The model of differential projection is primarily cognitive (see also Gramzow, Gaertner, & Sedikides, 2001; Otten, 2002). Although motivational variables may affect the strength of ingroup and outgroup projection, no particular motivational assumption needs to be made to explain the difference. The basic logic of induction is sufficient. As a rule, a sample of observations is most diagnostic of the population from which it is drawn. It may also be diagnostic of another population if there is reason to believe that population overlaps with the sampled population (Krueger & Acevedo, 2005). When there is no compelling reason, scientists and laypeople often resort to crude ordinal assumptions. Most believe, for example, that the behaviour of non-human primates tells them more about humans than does the behaviour of rodents: just how much more is hard to tell.

The first implication of the differential projection model is that the perception of intergroup differences does not require cognitive or motivational distortions. The meta-analytic effect sizes for ingroup and outgroup projection imply a correlation between ingroup and outgroup judgements of .05 (i.e., $.5 \times .1$). Because of differential projection, outgroup attributes may be seen as virtually independent of ingroup attributes. In contrast, if people projected equally to both groups, the resulting correlation would reveal moderate perceptions of similarity $.5 \times .5 = .25$. Empirical studies under the aegis of the common ingroup identity model exploit this relation to reduce intergroup discrimination (Gaertner, Mann, Murrell, & Dovidio, 2001).

Given that for real groups (e.g., women and men; Sunnis and Shiites) an all-inclusive category can always be found, the true between-group correlations over attributes can be expected to be positive. Therefore, the near-zero correlations arising from differential projection are most likely underestimates of similarity.[2]

The second implication is that projection minimises perceived intragroup differences. A person who is ignorant about the distribution of a certain attribute in the group may assume that each possible prevalence rate is equally likely. Aggregated over all these possibilities, the person's best estimate is that a specific group member has the attribute with a probability of .5. This estimate reflects maximum uncertainty and variability within the group. When projecting her own status with respect to the attribute (i.e, she does or does not possess it herself), she revises her estimate from .5 to 2/3 (Dawes, 1989; Krueger, 1998). The more people project, the more their probability estimates move towards 0 or 1, and at the limit, perceptions of within-group variability disappear. In as much as projection is stronger for ingroups than for outgroups, it follows that the former appear to be more homogeneous than the latter (Krueger et al., 2005). This analysis is supported in the minimal group paradigm, in which respondents do not have information about other group members and cannot apply social stereotypes. In real social groups, however, perceptions of outgroups are typically marked by homogeneity. These effects tend to be small (Mullen & Hu, 1989), but are noteworthy because they appear despite the countervailing effect of differential projection.

The third implication is that differential projection leads to ingroup bias if group members have positive self-images. The assumption of self-love is empirically sound. Hundreds of studies testify to people's willingness to endorse positive rather than negative attributes as part of their self-concepts (Alicke & Govorun, 2005). Ingroup judgements are not perfectly correlated with self-judgements, and are thus regressive with regard to the latter. If there is no other source of assumed ingroup positivity, the correlation between ingroup judgements and attribute desirability is the product of ingroup projection and self-love. Likewise, perceptions of outgroup positivity can be modelled as the product of outgroup projection and self-love. Because low outgroup projection yields even more regressive judgements, the resulting correlation is lower than the correlation involving ingroup judgements. The difference between the two correlations is a measure of ingroup bias.

[2]Of course, the true between-groups correlation could be zero, or even negative, for carefully selected attributes. However, representative sampling of attributes will likely yield a positive correlation.

The projection-based model of ingroup bias has several implications, which are supported by empirical research and presented here in list form.

1. Bias is reduced when self-judgements (and thus projection) are statistically controlled (Krueger et al., 2005; Otten & Wentura, 2001).
2. Bias is stronger among individuals with highly positive self-images than among individuals with neutral or negative self-images (Gramzow & Gaertner, 2005).
3. Bias decreases when outgroup members are recategorised as ingroup members (Gaertner et al., 2001).
4. Bias facilitates the rejection of atypical ingroup members and attraction to atypical outgroup members (Chen & Kenrick, 2002).
5. Bias is primarily a matter of favourable ingroup perceptions (Brewer, 1999). At least in the minimal group paradigm, outgroups are seen neither in a positive nor in a negative light. Derogation of real social outgroups implicates cognitive or motivational factors beyond the contributions of differential projection and self-love (Riketta, 2006).

The fourth implication of the differential-projection model is that greater projection is associated with greater accuracy. Lack of projection implies lack of accuracy, at least in the minimal group paradigm. This is a surprising result given the common view that, aside from outright outgroup derogation, perceptions of ingroups are positively inflated. The present model suggests that positive ingroup judgements need not be erroneous. Although the self-images of many individuals may be positively biased, self-love is a reliable psychological characteristic, and therefore ascriptions of positive attributes to others will be highly correlated with those others' self-judgements. In as much as an individual's profile of self-judgements is correlated with the profiles of others, the logic of induction ensures that judgements about others will be accurate in so far as they are projective. The correlation between an individual's self-judgements and the average self-judgements of other group members is an index of the degree to which the individual is a typical group member. Likewise, the correlation between the individual's self-judgements and his/her judgements of the group is an index of projection.

In as much as the typicality correlation is positive, a person will make more accurate judgements the more he/she projects. The correlation between group judgements and average self-judgements obtained in the group is an index of accuracy. Assuming that the individual has no other information, the accuracy correlation can be modelled as the product of the typicality and the projection correlations. In our studies with the crossed-categorisation paradigm projection predicted accuracy, and average accuracy was highest for the double ingroup, intermediate for the mixed group, and lowest for the double outgroup (DiDonato & Krueger, 2007).

In the minimal-group paradigm, or anywhere else where the individual has no information about others, projection cannot be exaggerated. However, there are other circumstances in which projection is inflated. Gilovich and his collaborators, for example, have conducted experiments showing that people overestimate how much attention others pay to them, how harshly they are judged by others after failure, and how privy others are to their own internal sensations. These phenomena, which are variously called the spotlight effect, the illusion of transparency, or empathy neglect appear to be cases of exaggerated projection.

Consider a study by Gilovich, Medvec, and Savitsky (2000). These authors asked participants to don an embarrassing T-shirt (featuring a picture of Barry Manilow), and to estimate how many others in an audience would notice it. As expected, participants overestimated the number of others who noticed the T-shirt. The difference between estimates and reality constituted *prima facie* evidence for projective bias. Notice, however, that the volunteers and the audience did not have the same phenomenal orientation and experience. The volunteers' internal perspective could not be treated as a representative sample of the perspectives available to the group. The audience was functionally an outgroup and the private information available to a volunteer was not projectible. If, however, participants were asked to rate their own discomfort and to estimate the discomfort of others *who are put in the same situation*, the accuracy benefit of projection could be modelled as inductive reasoning.

The spotlight effect is difficult to model because there is no opportunity to derive an optimal weight for self-information, and the effect is only demonstrated for individual prediction issues. Hence the participant's performance is compared with a standard that is difficult to meet. To avoid the charge of bias, the participant would have to make a spot-on prediction. Simply making a very low estimate will not do. In the T-shirt study, for example, predicting that no one noticed the embarrassing picture would have been a case of insufficient projection. To circumvent this difficulty, the induction model of projection recommends the use of multiple items. Then predictive accuracy can be measured as the association between predictions and social reality.

An alternative strategy is to devise a prediction task in which the normative effect of self-knowledge is in fact zero. Van Boven and Loewenstein (2005) found, that people project whatever transient need states they experience at the time onto others. Participants who were made to feel thirsty in the lab were more likely to attribute thirst than hunger to a hypothetical hiker lost in the woods, whereas participants whose thirst was not aroused did not show this effect. Arguably, both hunger and thirst are problems in the outdoors, and an indoors perceiver's momentary experience of either one of these needs has no cue validity for predictions.

Summary and review

Social projection ranks among the most robust and replicable phenomena in social-perception research. The evidence suggests that projection is a primitive judgemental heuristic that is easily and preconsciously engaged, but whose operation can be modulated by deliberative thought. Social projection can be distinguished from self-stereotyping, which, when it occurs, also yields positive correlations between self-judgements and group judgements. On balance, however, self-stereotyping appears to be the more fragile phenomenon. In intergroup contexts, differential projection contributes to perceptions of intergroup differences, ingroup homogeneity, ingroup favouritism, and comparatively inaccurate judgements of outgroups. These contributions are most clearly seen in minimal laboratory situations, where pre-existing social stereotypes play no role. Still, the effects of social projection can help to identify baseline expectations for judgements of real social groups, so that phenomena such as perceived outgroup homogeneity or outgroup derogation may be better understood.

SOCIAL PROJECTION IN SOCIAL DILEMMAS

In the preceding section, social projection was considered in contexts in which a person has access to the components of a rather stable self-concept. When making predictions about others, the perceiver can simply look up his/her own attributes and use them as projectible cues. The situation is different when there is no self-concept that can be looked up. What if people's decisions depend in part on what others do, or on what they think others will do? Social dilemmas present this type of problem. In a social dilemma, a person's outcome depends not only what he/she chooses to do, but also on what others do. I begin by discussing the simple case in which the person has chosen, for whatever reason, one of the two possible courses of action. This choice is a projectible event. Then I consider social dilemmas more fully by asking how projection can influence a person's choice in the first place. Here, I will outline a theory of "evidential decision making" (Jeffrey, 1964/1983), and review recent findings that support it. Finally, I extend this discussion to situations that comprise both an intra-group and an intergroup dilemma. My interest lies in elucidating the role of differential projection (i.e., a pattern of strong projection to ingroups and weak projection to outgroups) in shaping individual and collective outcomes.

Projection after choice

The most famous social dilemma is the prisoner's dilemma (PD; Poundstone, 1992). In the PD, the interests of the individual collide with

the interests of the collective. According to the canonical story, two suspects are apprehended and charged with a felony. The prosecutor, who needs at least one confession for a conviction, talks to each suspect independently and offers the following deal: "If you confess and the other does not, you get off on probation, while your accomplice will serve 9 years in prison. If you both confess, you will each serve 6 years. If neither one of you confesses, you will both be convicted on a lesser charge and serve 3 years."

Classic game theory holds that a rational player defects because a confession yields a shorter sentence regardless of what the other player does. Mutual defection is an equilibrium because neither player can do better by switching to cooperation. This equilibrium is not efficient, however, because both players could do better if they found a way to cooperate with each other. In an anonymous one-shot dilemma, there is no opportunity to negotiate, make promises, or appeal to feelings. Even if the players were to communicate, the structure of the game would be the same because both would face the temptation to renege on a promise.

The empirical data do not support classic game theory. In a typical study, close to 50% of players cooperate (Sally, 1995). Setting aside for a moment the question of how they arrive at their choice, it can be noted that after choice, social projection operates much as it does in other prediction contexts. On average, cooperators believe that the probability of others cooperating is between .6 and .7; and defectors expect defection from others with a similar probability (Dawes et al., 1977; Deutsch, 1960; Messé & Sivacek, 1979).

It is unlikely that players generate predictions first and then choose for themselves. If that were so, it would be necessary to explain why some people expect cooperation to be more probable than defection, whereas others expect the opposite in the first place. The one-shot anonymous PD is much like the minimal group paradigm in that it excludes specific information about individual others and general stereotypes about groups. Dawes et al. suspected that players have pre-existing preferences, which they enact in the game and then project accordingly. To test this hypothesis, Dawes et al. also asked observers to make predictions about players' choices. As expected, players' predictions were not only correlated with the players' own choices, but they were also more variable than the observers' predictions. This finding suggested that the players' choices contributed systematic variance to their predictions (see the earlier section on projection vs self-stereotyping). According to the conformity or "bandwagon" hypothesis, the distribution of predictions should have been the same for players and observers.

Now consider the PD more formally, using the following notation: T is the Temptation payoff obtained from unilateral defection; R is the Reward payoff obtained from mutual cooperation; P is the penalty payoff obtained

from mutual defection; S is the sucker's payoff obtained from unilateral cooperation (Rapoport, 1967). By the classic analysis, defection dominates because a defector is better off than a cooperator regardless of what the other player does. A player only needs to be able to subtract to see that $T - R > 0$ and that $P - S > 0$. The estimated probability of the other player cooperating, p_c, is irrelevant. Whatever the value of p_c is, the expected value of cooperation, EV_d, is greater than the expected value of defection, EV_c.

If classic game theory did not ignore p_c, it would still insist that its value must be fixed for any individual player. Once a player has estimated p_c, he/she can formulate his/her own choice. Unfortunately, classic game theory implies that making the expected p_c explicit has a self-eliminating effect. Whatever the value of p_c might be, rational players will defect and thereby negate the truth of their expectations. In an experimental test of this idea, participants received the values of .5, .75, or 1 as the probability with which a computerised player would make the cooperative choice. In each condition, about four out of five participants defected, suggesting that they knew how to maximise their own payoffs (Krueger & Acevedo, 2005, experiment 2; see also Shafir & Tversky, 1992).

Cooperators who project their own choices onto others cannot do this with the aim of maximising their expected values. Nevertheless, projection appears to be rational in that it maximises the expected accuracy of prediction. Recall that under the assumption of zero information (i.e., "ignorance" or "uniform priors"), the optimal prediction is that two out of three others will choose as one does oneself. In the PD, accuracy is $p_c^2 + (1 - p_c)^2$ for a player adopting a strategy of probability matching. If a player believes, for example, that others cooperate with a probability of two out of three and cooperates with the same probability, the probability of a correct prediction is 5/9, or .556. The player would do better by always cooperating so that the probability of being correct reaches its ceiling of p_c. The conclusion that projective predictions maximise accuracy as long as the self is a valid cue for the group restates the conclusion reached earlier in the area of projection over multiple personal attributes. There, a perfect projection correlation over attributes sets the accuracy correlation to its maximum, which is the size of the typicality correlation (as long as the latter is larger than zero).

Although post-choice predictions in the PD can be modelled as Bayesian inferences, an alternative explanation must be considered. Consider a defector who expects cooperation from others. This player would appear to be an unabashed exploiter. In contrast, a cooperator who expects defection would appear to be a simpleton (Krueger & DiDonato, 2007). To avoid negative implications for self-perception and reputation, players rather have to make predictions consistent with their choices. These concerns do not apply to projection outside dilemma situations. In the classic study on

consensus bias, for example, participants either volunteered for a second study or they declined, and most of them thought that most others would choose as they themselves did (Ross, Greene, & House, 1977). In this case, the deniers who expect that a majority will volunteer are not seen as particularly selfish, and volunteers who expect that the majority will decline are not seen as particularly dull (Krueger & DiDonato, 2007).

Projection before choice

If expectations regarding the choices of others can be modelled as projective inferences drawn from players' own choices, the question remains how these choices were made in the first place. In its anonymous one-shot realisation, the PD offers few cues players can use to formulate preferences. Yet it is precisely this paucity of information that makes projection possible at the pre-choice stage. Here, the projection hypothesis becomes an instance of evidential decision theory (Jeffrey, 1964/1983), which suggests that people can generate expectations about what others might do based on their own, potentially variable, inclinations. Indeed, the defining characteristic of the PD is that players still need to decide. They have to puzzle over how their own choices will combine with the choices of others to yield final outcomes. This puzzling can entail the consideration of different scenarios. Specifically, players can entertain two possible states of the world depending on their own, yet-to-be realised, behaviour. They may wonder, "What would I predict the other player will do if I were to cooperate? What would I predict if I were to defect?" If ignorance is complete with regard to the other, the optimal Bayesian inference is that the other player will choose whatever the predictor chooses with a probability of 2/3. This probability will henceforth be written as p_r, where the subscript stands for reciprocation.

The player can then use these two conditional probabilities to assess the expected value of the game. The expected value of cooperation is $p_rR + (1 - p_r)S$, and the expected value of defection is $p_rP + (1 - p_r)T$. If, for example, $T = 4$, $R = 3$, $P = 1$, $S = 0$, then $EV_c > EV_d$ if $p_r > \frac{2}{3}$. It may not be assumed that all players begin with the assumption of uniform priors or that projection is equally strong for all. In as much as there is variability in the perceived probability of reciprocation, some players will conclude that cooperation maximises own payoff, whereas others will not. In theory, however, each player's "choice is based on which option confirms the best kind of news" (Levi, 2000, p. 390).

Value maximisation takes the details of the payoff matrix into account. Rapoport (1967) suggested $\frac{R-P}{T-S} = K$ as an index of matrix difficulty. A smaller ratio means that it is more difficult for a player to cooperate. From the perspective of classic game theory, the designation of a matrix as difficult is meaningless because players should always defect. However, if social

projection comes into play, the specific payoffs are critical. They yield the probability of reciprocation at which cooperation becomes attractive. Specifically, $EV_c > EV_d$ if $p_r > \frac{1}{1+K}$ (Acevedo & Krueger, 2005; Brams, 1975).

We designed an experiment to see if people are sensitive to variations in expected values. To see if they can make the necessary computations (or recruit intuitions that closely mirror them), participants were informed that the other player was a computer that was programmed to reciprocate whichever option they chose with a probability of .5, .75, or 1. As expected, the rate of cooperation increased sharply across these levels of p_r (Acevedo & Krueger, 2005, experiment 1). Also consistent with the model, rates of cooperation depended on matrix difficulty when the probability of reciprocation was intermediate.

A possible concern about this experimental design is that players were paired with a computer instead of a human being, and that they perceived their task as a technical matter of numerical estimation. We therefore conducted an experiment in which we subjected the projection hypothesis to a particularly difficult test; that is, we assumed that the likelihood of cooperation is predictable from the strength of a person's pre-existing tendency to project. People with a strong tendency to project should transfer expectations of similarity to the PD, and thus be more likely to expect the rewards of mutual cooperation than people who are less inclined to project. After assessing each respondent's projective tendency idiographically by correlating self-judgements with judgements of others across various personality descriptors, we found a significant correlation (point-biserial $r = .15$) between the strength of projection and the choice of cooperation over defection (Krueger & Acevedo, in press, experiment 3).

For most versions of the PD, it is true that the more people project—and the more strongly they project on average—the more efficient is the collective outcome of the game. Efficiency can be defined as the sum of the payoffs extracted from the experimenter. The straight route from projection to cooperation to efficiency holds for the most commonly researched type of PD, namely a game with a balanced payoff structure. A payoff structure is balanced if the difference between the Temptation payoff and the Reward payoff is the same as the difference between the Penalty payoff and the Sucker's payoff. When $T - R = P - S$, each cooperative choice increases efficiency because the amount of money sacrificed by the cooperator is smaller than the amount gained by the defector. Over repeated plays, total efficiency increases with the rate of cooperation. That is, efficiency is maximal for $p_c = 1$.

However, the payoff structure of a game may be unbalanced (Coombs, 1973; Dawes, Orbell, Simmons, & van de Kragt, 1986). If $T - R < P - S$, the game stimulates fear. Although full cooperation is still the most efficient

strategy, the relationship between cooperation and efficiency is no longer monotonic. If S is much smaller than the other three payoffs (i.e., if $2P > T + S$), certain cooperation is only slightly better than certain defection, and both are more efficient than a mixed strategy of cooperating some of the time or with an intermediate probability. This curvilinear relationship between p_c and efficiency means that the relationship between projection and efficiency is also curvilinear. Efficiency would be highest if projection were so strong that everyone cooperated or so weak that hardly anyone cooperated. A fear-inducing PD with a strong payoff imbalance would be very difficult (i.e., have a low K value), which suggests that defection would be common and the second most efficient outcome would be obtained.

The opposite type of structural imbalance occurs when a payoff matrix stimulates greed. When $T - R > P - S$, the Temptation payoff is the outlier. If the inequality is large (i.e., if $2R < T + S$), then efficiency does not increase monotonically with the probability of cooperation. Instead, the relationship is an inverted U-function where the value of p_c that maximises efficiency converges on .5 as the payoff imbalance becomes more extreme. As the matrix becomes more difficult, projection is decreasingly able to induce cooperation, and p_c falls. As a result, the least efficient outcome of mutual defection is likely obtained.

To review, pre-choice projection amounts to an increase in the perceived probability of bilateral outcomes. When projection is strong enough, players can choose to cooperate in an effort to maximise the expected value of the game. Projection thereby disables the dominance principle, which mandates defection. When the payoff structure is balanced, which it is in most research studies, projection also increases the efficiency of the collective outcome.[3]

Other games

Pre-choice projection encourages cooperation in any game as long as the mutual cooperation payoff is larger than the mutual defection payoff. To illustrate, consider the games of chicken, assurance, and coordination. The game of chicken is similar to the PD, except that mutual defection is catastrophic (i.e., $T > R > S > P$). As such, this game is a formal analogue of an arms race (Russell, 1959). Although defection is not dominating, there is no guarantee that mutual cooperation will be achieved. Each player is tempted to defect if he/she suspects the other will cooperate. As in the PD, social projection guards against this temptation. As soon as a player contemplates switching from cooperation to defection, the projected choice

[3]It is unknown, and perhaps unknowable, how typical balanced payoff structures are in the social world relative to fear- or greed-biased games.

of the other switches too, making the catastrophic P outcome more likely. Compared with its role in the PD, social projection is even more effective in the game of chicken because for a given set of payoffs, this game has a higher K value. For example, $K = .5$ in the PD implies that $K = 1$ in the game of chicken. Thus, a lower level of projection is sufficient to induce cooperation.[4]

In an assurance game (also called "stag hunt") the mutual cooperation payoff is higher than the unilateral defection payoff ($R > T > P > S$). Still, there is no guarantee of cooperation. A player contemplating cooperation might switch to defection because he/she fears that others do too. This line of thought signals a bandwagon effect, which social projection can hold in check. As the player switches back to cooperation, his/her expectation that others will also cooperate ought to be strengthened. As in the game of chicken, even a weak tendency to project is enough to induce cooperation.

A coordination game presents a greater departure from the PD. Here, the unilateral defection payoff drops to the third rank ($R > P > T > S$). Although there is no guarantee of cooperation, most naïve players regard this game as trivial. Schelling (1960) noted that classic game theory cannot explain how players manage to converge on mutual cooperation so easily. From the perspective of evidential decision making, however, there is no mystery. A modicum of pre-choice projection suggests that cooperation maximises the value of the game.

Social preferences

The social projection hypothesis suggests that players who choose cooperation over defection can do so for self-regarding reasons alone. The increase in the overall efficiency of the game, as reflected in the collective's take, can arise as a by-product of this individual decision strategy. This idea is theoretically important because it explains cooperation without imputing social preferences to individual agents. In contrast, theories of social preferences assume that people choose cooperation only if they care enough about the welfare of others or about the allocations being fair.

Van Lange (1999, 2000) proposed that people transform a given payoff matrix into a subjective one by weighting the payoffs for others and the differences between their own and the other's payoffs. The effective payoffs are modelled as EP = own payoffs + $w_1$other's payoff − w_2|own

[4]Researchers typically do not ask how to achieve mutual cooperation in the game of chicken, but how to get the other player to chicken out. The methods needed for success require communication. For example, a player who manages to convince the other that he/she is crazy enough to defect will force the other to cooperate (Schelling, 1960).

payoff – other's payoff|. A player who is benevolent but not averse to inequality (i.e., $w_1 > 0$ and $w_2 = 0$) finds that cooperation dominates defection (i.e., $R > T > S > P$) if $w_1 > \frac{T-R}{R-S}$ and $w_1 > \frac{P-S}{T-P}$. If the given matrix is balanced, the two ratios are the same. If the matrix is unbalanced, w_1 must exceed the larger of the two ratios. A positive weight for fairness reduces the effective T payoff, but it also reduces the S payoff. The surprising result is that a preference for equality keeps cooperation from becoming the dominating choice. A desire for fairness only transforms the PD into a stag hunt, and if the desire is strong, into a coordination game. The question of how players come to cooperate is thus postponed rather than solved. Social projection remains relevant. If social preferences are at play, the effective games are easier than the objective games, and therefore a lower level of projection is sufficient to induce cooperation.

Objections and rebuttals

Philosophers and decision theorists continue to debate the status of evidential decision theory (Joyce, 1999). I now present two commonly raised objections against this theory and try to disarm them. The first objection is that individuals ought not to generate different expectancies depending on hypothetical differences in their own behaviour. There is, after all, only one true value of p_c, and therefore, a person who generates two different values will surely make an error. This objection is the same that was levelled against Dawes's (1989) Bayesian interpretation of consensus bias. However, Dawes showed that individuals with different sample information (i.e., their own behaviours) can minimise their aggregated errors by making different predictions. If everyone guessed consensus while ignoring own behaviour, the sum of the errors would be larger. This analysis was ultimately accepted when it was realised that each individual made an optimum prediction given the information they had.

When the same person makes different predictions based on behaviours that have not yet occurred (and only one of which can ultimately occur), the Bayesian rationale is the same. To illustrate the statistical identity of real and hypothetical behaviours, consider again the player's line of reasoning. He or she can infer that "Once I have cooperated, my best estimate for the probability of cooperation will be 2/3. My choice will be more likely the choice of the majority than the choice of the minority. Alternatively, once I have defected, my best estimate for the probability of cooperation will be 1/3." The distinction between a person's own future, yet unrealised, behaviour, and his/her past behaviour is spurious from a statistical point of view. If one accepts the Bayesian rationale for post-choice projection, one must also accept it for pre-choice projection. Not to do so would be incoherent, and thus irrational.

Proponents of "causal decision theory" reject this view, arguing that the choices of two individuals are conditionally independent once common causes are removed (Eells, 1991). That player B is more likely to choose like player A rather than choose differently can be attributed to a common cause, C, which "screens off" the effect of A on B (and likewise the effect of B on A; Reichenbach, 1956). When $p(A\&B|C) = p(A|C)p(B|C)$, it is also true that $p(A|B\&C) = p(A|-B\&C)$. What does this mean for the decision maker? The decision maker concludes that either C or $-C$ must be true. Thus, whichever choice he/she makes, there are no implications for what the other player does. It is therefore best to defect.

The problem with this objection is that it assumes something or someone that holds C constant. That works fine in theory, but it is not helpful to the decision maker in the PD. Without a credible assurance that C or $-C$ has been fixed, the decision maker returns to the idea that his/her own choice is diagnostic. Now it is diagnostic of C, which in turn is diagnostic of the other player's choice. The end result is the same; the only difference is that the mediational path from A to B via C has replaced the direct path from A to B. To block evidential inferences from A to C, the theorist has to invoke another cause D that affects both A and C, and repeat the argument, ad infinitum.

The second, and more critical, objection is that one should not use differential conditional expectancies to make decisions, however legitimately these expectancies may have been generated. Players who cooperate *because* cooperation yields a higher expected value than defection are charged with "magical thinking" (Quattrone & Tversky, 1984). Magical thinking here stands for magical causation. Clearly, individual players cannot *make* others cooperate simply by cooperating themselves. Yet, according to this criticism, this is what a value-maximising player appears to be trying to do. By choosing to cooperate, one player can infer that the other player will probably cooperate too. Had the first player chosen to defect, he/she would have inferred that the other player probably defected too. As the first player's decision to cooperate or to defect is cast as a deliberate choice, it seems that he/she claimed to have causal power over the choice of the other. If such power were granted to one player, it would diminish the power of the other. The PD cannot grant a power differential, however, because neither player holds a privileged status.

The charge of magical causation is that the player seeks to control the other by cooperating. The charge of "seeking to control the other" implies that the player is seen as an agent with intentions and the ability to act differently. If the player were to defect and claim he/she does not care what the other does, the player would be free from the charge of magical causation. Suppose, however, that the player's "decision" were understood as the outcome of automatic processes rather than the result of a deliberate

or "free" choice. Upon seeing the payoff matrix, a player senses an urge to defect. As time passes, the player's experienced intention switches to cooperation. Back and forth shifting continues much like the switching of visual perception between different representations of a Necker cube (Attneave, 1968). Ultimately, one of the two possible frames is retained, or "chosen", if for no other reason than that time was up. The final decision might as well be fully determined by nonconscious processes and it might as well be cooperation. Whatever his/her final response is guides the player's expectation regarding the other's final response.

By referring only to the statistical dependencies among individuals' actions, evidential decision theory makes no distinction between freely chosen and determined responses. A player might *feel* that his/her choice was freely willed, and this feeling can be taken as an instance of magical causation (Pronin, Wegner, McCarthy, & Rodriguez, 2006). However mistaken they might be, however, beliefs regarding the causal genesis of choice have no bearing on the rationality of the choice itself. Ultimately, players' choices are mutually predictable in the statistical sense without being mutually caused (Brams, 1975).[5,6]

Players who take the charge of magical causation seriously might wonder what they should do instead. As choosing randomly hardly seems rational, only the classic mandate of defection remains. Defectors may have the satisfaction of meeting the demands of traditional rationality, but do poorly as a group. Once they defect, they should not believe that the other probably cooperated. Evidential decision making needs to be either employed rigorously or not at all. Consider the futility of trying to cheat the logic of induction. Suppose a player contemplated cooperation, with the attendant expectation that the other player would probably cooperate too. Then the player ever so rapidly switches to defection while freezing the expectation of cooperation. Now *this* manoeuvre is a good example of magical thinking. The inductive implications of self-generated evidence cannot be outrun, much as it is impossible to look into the rear view mirror fast enough to catch oneself with one's eyes still on the road.[7]

To test whether people are sensitive to the implications of switching strategies, we conducted experiments in which participants read about four types of player. Two players were described as considering cooperation and

[5]Mutual predictability does not entail a positive correlation between choices across pairs of players. It only entails that the probability of a matching choice is higher than the probability of a mismatching choice.

[6]During his sceptical period, Bertrand Russell rejected the notion of causality as flowing forward through time. Determinism "makes no difference between past and future: the future 'determines' the past in exactly the same sense in which the past "determines" the future. The word 'determine,' here, has a purely logical significance" (Russell, 1932, p. 195).

[7]I thank to Lauren Krueger for this simile.

as believing that the other player would probably cooperate too. Two other players were described as considering defection and believing that the other player would probably defect too. Each player was then offered a last-moment opportunity to switch. The switch was either unilateral, meaning that the other player's choice remained unchanged, or bilateral, meaning that whatever the other player had chosen would be inverted as well. If participants used the players' expectations to maximise the expected value of choice, they would advise a cooperator to switch to defection if the switch could be made unilaterally, and they would advise a defector to switch to cooperation if the switch could be made bilaterally. Indeed, this is what most participants did. In a second experiment, no expectations were provided. Instead, participants made their own estimates regarding the expectations held by a presumed cooperator and a presumed defector. Still, the same pattern of recommendations was observed because participants reasonably assumed that players would project (Krueger & DiDonato, 2007).

Intergroup discrimination in dilemmas

The preceding analyses have shown that, for individuals with positive self-images, differential projection can yield ingroup favouritism and increased cooperation with other ingroup members. Now the question is how these effects combine when individuals make choices in an intergroup context. Can differential projection explain why people often discriminate behaviourally against outgroups? If so, intergroup discrimination may be, at least in part, traced to heuristics of inductive reasoning and the goal of maximising individual payoffs.

The central assumption of the social projection hypothesis is that group members believe that their own behavioural choices forecast what other ingroup members will do while remaining uncertain about the behaviour of outgroup members. Models of (perceived) behavioural interdependence offer similar accounts for reward allocations in the minimal group paradigm (Rabbie, Schot, & Visser, 1989). Participants typically distribute points preferentially to ingroup members, but they do so only when they believe that other ingroup members also perform the same allocation task (Gaertner & Insko, 2000; Stroebe, Lodewijkx & Spears, 2005). Benefiting an anonymous ingroup member creates an expectation of generalised reciprocity. If not this ingroup member, then perhaps some other member will benefit the self. Brewer (in press) has termed this heuristic "depersonalised trust". Yet the concept of depersonalised trust does identify differential projection as the critical operative mechanism. Behavioural ingroup favouritism could occur simply because people hold the expectation that, compared with outgroup members, ingroup members have stronger

social preferences to benefit members of their own group (i.e., a high value of p_c; Yamagishi & Kiyonari, 2000). If so, ingroup favouritism would be an instance of conformity to a perceived behavioural norm; it would not answer the question of how this norm arose in the first place.

The social projection hypothesis can readily be applied to intergroup conflicts. Individuals can ask themselves what their own behaviour implies about the behaviour of other ingroup members and outgroup members. In as much as they project their own behaviours more strongly to ingroups than to outgroups, people may conclude that they personally fare better if they discriminate. Failing to project to outgroups is risky. On the one hand, people might overlook the inclination of outgroup members to favour members of their own group. On the other hand, the attribution of ingroup favouritism to the outgroup could escalate the conflict by supporting further increases in the ingroup's own favouritism (Krueger, 1996). Projection to the outgroup could help people to expect egalitarian behaviour from members of any group, but only if they themselves are prepared to engage in egalitarian behaviour. In this section, I explore the implications of differential projection for intergroup discrimination in the context of nested social dilemmas.

Nested social dilemmas. When an intergroup dilemma is superimposed on an interpersonal dilemma, the resulting game is a model of intergroup conflict (Kahan, 1974; Schlenker & Bonoma, 1978). Consider the most extreme case, war: "The tension between group welfare and individual welfare is starkest when group solidarity entails risking one's life: All members benefit if the group acts collectively in defense of shared interests, but even moderately sensible members might hesitate before joining a possibly fatal fray" (Gould, 1999, p. 359). In war, an individual's strategy not to cooperate with other group members literally amounts to defection. The defector seeks to take a free ride on the contributions of others to the group effort.

Note that Gould's description only refers to the interpersonal dilemma, where individual self-interest conflicts with the interest of the ingroup. Yet the intergroup conflict can only be understood by considering the joint outcomes arising from the four combinations of majority behaviour in both groups. If most individuals in both groups cooperate by risking their lives, the collective outcome is mutual destruction. However, if most members of one group cooperate, whereas only few members of the other group cooperate, the first group will defeat the second group. From the perspective of the overall collective, it would be most efficient if most members of both groups defected, in which case the intergroup outcome would be peace. What if, as Carl Sandburg asked, there was a war and no one came?

In a nested dilemma, an individual has to contemplate 16 possible payoffs. In the sample display (see Figure 1), each payoff is the result of the individual's own choice, the choice of the ingroup majority, and the choice of the outgroup majority. Each of the four quadrants of the full matrix is a rather difficult interpersonal dilemma. The intergroup dilemma can be seen by considering the average payoffs of the four quadrants. If both groups cooperate, the average payoff is 5; if both defect, the average is 7.5. If one cooperates while the other defects, the respective payoffs are 10 and 2.5. The best outcome for the group is when it achieves a high rate of cooperation, while the other group does not (see upper right matrix). The second best outcome is when both groups have a high rate of defection (lower right). The third best outcome is when both groups have a high rate of cooperation (upper left). Finally, the worst outcome for the group is when only the other group has a high rate of cooperation.

From the group's perspective, cooperation is the dominating choice. No matter how the majority of the outgroup decides, the collective ingroup payoff is higher if most ingroup members cooperate rather than defect. If the group were a classically rational agent with a mind of its own, it would seek to satisfy its self-interest by generating a high rate of cooperation among its members. In times of intergroup conflict, real social groups do just that by using the full arsenal of influence techniques. Political propaganda seeks to shape patriotism and it promises spoils of war, while also making sure that individuals remember the personal risks they would face if they were to defect (Coser, 1956; Stouten, De Cremer, & van Dijk, 2006). In other words, it falls to the powerful élites to "play the game of war" by trying to anticipate the moves of other élites and to control the "choices" of their own group members. To quote Gould (1999, p. 258) again, "The Hobbesian problem of conflict between groups arises only because it has been solved within groups."

More optimistically, Lodewijkx, Rabbie, and Visser (in press) suggest that small, unstratified groups are neither more nor less rational than individuals. At least when it is known that the intergroup games will be repeated over a number of rounds, group members tend to realise the benefits of "cautious reciprocation", and thereby avoid mutually destructive outcomes (see also Wildschut & Insko, 2006). Collectively desirable outcome are further enhanced by discussions involving members of both groups (Bornstein, 2003).

Differential projection stimulates conflict. Like the standard two-person prisoner's dilemma, the nested social dilemma is most difficult to solve when it is presented in its anonymous one-shot form. Consider a case in which ingroup projection is strong enough to make cooperation attractive ($p_r = .8$). Further assume that there is no outgroup projection ($q_r = .5$). Individuals

believe that it is equally likely for the outgroup majority to match the behaviour of the ingroup majority or to act differently.[8] When considering cooperation, a player assesses the expected value of cooperation using the payoffs in the top row of Figure 1, calculating that $EV_c = .5(.8 \times 6 + .2 \times 2) + .5(.8 \times 12 + .2 \times 4) = 7.8$. The values in the second row are now irrelevant because they refer to a person who considers defection while believing that the ingroup majority cooperates. When considering defection, a player assesses the expected value of defection using the payoffs in the bottom row of the Figure, calculating that $EV_d = .5(.2 \times 4 + .8 \times 2) + .5(.2 \times 12 + .8 \times 6) = 4.8$. Here, the third row is moot because it refers to a cooperator who expects the ingroup majority to defect. The difference in the expected values suggests that a self-regarding player will cooperate rather then defect. Not knowing what the outgroup will do, the person wonders whether there will

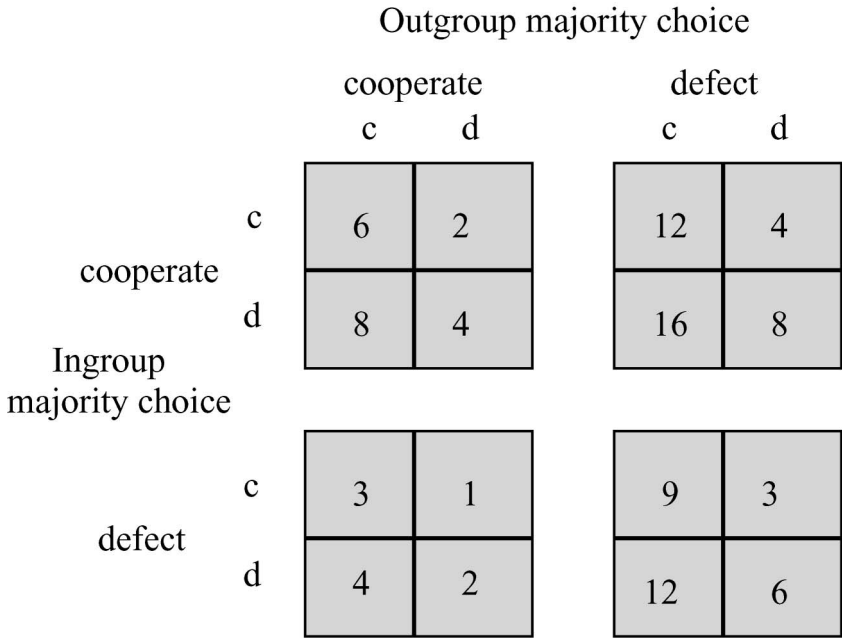

Figure 1. Payoff matrix for a nested prisoner's dilemma. *Note*: "c" and "d" respectively signify cooperation and defection by an individual player.

[8]The value .5 for q_r is arbitrary. What matters is only that the same probability is used for both parts of the expression.

be mutual destruction or victory. In contrast, an individual who chooses defection as the strategy that is dominating at the individual level will wonder whether there will be defeat or peace.

From a bird's eye perspective, collective welfare (i.e., peace) is the most desirable outcome, and one can ask whether it is possible to increase it by reducing the ingroup – outgroup differential in social projection. When people are categorised into minimal groups, their ingroup projection is resistant to change. Outgroup projection, however, becomes stronger when participants first make social predictions for the superordinate population. In that situation, they appear to realise that the outgroup is, after all, subsumed under the same collective that also contains the ingroup. In contrast, social categorisation is more salient when no such prior consideration of the superordinate collective occurs. Then, people hardly project to the outgroup at all (Krueger & Clement, 1996). As noted above, any reduction in the projection differential results in a reduction of perceptual ingroup favouritism. The question is whether the same is true for behavioural ingroup favouritism.

Intuitively, it may seem that if people perceive all others as members of a superordinate ingroup, and if they project accordingly, while ignoring group boundaries, the intergroup social dilemma will dissolve. Suppose a player not only believes that ingroup members will choose as he himself does with $p_r = .8$, but also that the majority choice of the outgroup will likely be the same as the majority choice of the ingoup (i.e., $q_r = .8$). Now, the expected values for cooperation and defection are the same. This is a general result: $EV_c = EV_d$ if $p_r = q_r$. As soon as a person projects however slightly more within the group than across groups, $EV_c > EV_d$, thus enabling collectively undesirable outcomes. Consider two illustrative cases. In the first case, where $p_r = .9$ and $q_r = .85$, the result is that $EV_c = 6.44$ and $EV_d = 5.94$. In the second case, where $p_r = .58$ and $q_r = .53$, the result is that $EV_c = 6.35$ and $EV_d = 5.85$. The value-maximiser cooperates in both cases, and does so whenever $p_r > q_r$. In short, attempts to solve the nested dilemma with appeals to a common, superordinate identity can succeed only if that identity is fully accepted. Partial acceptance yields the same outcome as no acceptance at all.

Groups are more competitive than individuals. In the interpersonal prisoner's dilemma, projection can induce individuals to cooperate and thereby contribute to the common good. However, the degree of projection necessary to make cooperation attractive tends to be high, with its minimum depending on the difficulty of the payoff matrix. Conversely, no particular level of projection in the nested dilemma draws a value-maximising player towards defection. As long as ingroup projection exceeds cross-group projection, rational maximisers will want to cooperate and thereby act

against the interest of the collective. A modest degree of ingroup projection is sufficient grounds for cooperation, as long as projection across groups is even lower.

This analytical finding reflects the well-established interpersonal–intergroup discontinuity effect, whereby groups are considerably more competitive with one another than individuals are. The standard explanation of this effect involves a family of motivational factors that enhance group members' greed for temptation payoffs and their fear of being suckered by the outgroup (Wildschut, Pinter, Vevea, Insko, & Schopler, 2003). Evidential decision theory suggests that differential projection can produce this discontinuity even when the motives of fear and greed are not at play.

Consider the empirical finding that the discontinuity diminishes as payoff matrices become easier.[9] Recall that in the interpersonal game cooperation depends on whether projection is stronger than the threshold set by the difficulty of the matrix. The more difficult the matrix is, the fewer players will cooperate. In contrast, cooperation with the ingroup in an intergroup game, which amounts to competition with the outgroup, only requires that ingroup projection exceeds cross-group projection. Matrix difficulty is irrelevant in the intergroup context. In other words, to understand why harder matrices yield larger interpersonal–intergroup discontinuities, it is enough to know that such matrices make it harder for individuals to cooperate in purely interpersonal games.

To be sure, the finding that differential projection can yield an individual–group difference in competition does not mean that motivational forces are irrelevant. Compelling evidence for the possibility of greed to override the effects of social projection comes from a recent study by Insko, Kirchner, Pinter, Efaw, and Wildschut (2005, experiment 2). In addition to the payoffs characteristic of the prisoner's dilemma, participants had an "exit" option guaranteeing a payoff "E" that was intermediate in value and independent of the choices made by others (i.e., $T > R > E > P > S$). The critical finding was that spatially segregated groups were more likely to compete with each other when they were categorised as members of the *same* psychological group (e.g., people who preferred Klee over Kandinsky paintings) than when they were categorised as members of different psychological groups. The latter were most likely to exercise the exit option. This pattern suggests that members of spatially different, but

[9]Insko and colleagues refer to the "noncorrespondence of payoffs", which is computed from the correlation between players' (or groups') payoffs across the four possible outcomes of the game. After Fisher Z transformation, the correlation between the index of non-correspondence and the K index of matrix difficulty is .72 across all possible PD games using $T = 12$ and $S = 0$ as their anchoring payoffs.

psychologically similar, groups yielded to greed, believing that the outgroup was exploitable. According to the social projection model, only cooperation or competition can yield the maximum value.

Conflict reduction. Efforts to overcome intergroup conflict in a nested dilemma can take various forms. One could try to reduce ingroup projection, increase outgroup projection, or bypass projection altogether. In a one-shot nested game, reducing ingroup projection to the point that it is nogreater than outgroup projection is difficult. A reduction of ingroup projection would have no effect on final outcomes unless that reduction completely eliminated the ingroup – outgroup differential. Ingroup projection is partly automatic and lowering it would diminish the accuracy of prediction. Indeed, ''debiasing'' studies have not been very successful (Alicke & Largo, 1995; Krueger & Clement, 1994). Likewise, it is difficult to see how an individual might come to project more strongly to the outgroup than to the ingroup, yet this would be the prediction strategy necessary for defection to become attractive.

If individuals are primarily self-interested (Pruitt & Kimmel, 1977), and if local élites are concerned with the success of their own group at the expense of others, who is to champion the interests of the collective? Indeed, social élites will rather try to foster positive expectations of reciprocity within the group, and, if possible, negative expectations with regard to the outgroup.[10] Therefore, it would seem tempting to return individuals to the type of rationality envisioned by classic game theory. A classically rational person could ignore the behaviour of outgroups, and simply defect in the interpersonal game. The benefit to the collective would be an unintended by-product of this strategy.

Nested social dilemmas are among the most risk-fraught and the least tractable social phenomena, and morally sensitive people are bound to be confused. Having learned to identify the demands of moral norms with self-restraint, they now realise that these norms are not concerned with the universal welfare of humanity, but only with the ingroup's competitive advantage over an outgroup. The paradox arising from this layered morality is that narrowly defined individual self-interest is consistent with the welfare of the collective. The welfare of the parochial group is caught in the middle, and inconsistent with both.

The nested prisoner's dilemma highlights the context-dependency of certain value judgements. When cooperation is seen as a choice that provides the good of the group, it appears to be desirable. However, when the welfare of the superordinate collective is considered, cooperation with

[10]Savvy élites induce ordinary people to make sacrifices for the group that they themselves would not dream of making.

the ingroup appears in a less favourable light. Likewise, the heuristic of differential projection now appears to be a damaging way of thinking. When social projection cannot be relied on to solve nested social dilemmas, contracts and enforceable agreements at the intergroup level may just remain the best hopes for mutually efficient outcomes because they do not depend on decisions at the individual level.

To complicate matters further, some intergroup competition is beneficial to collective welfare. At least at the level of small groups with conflicting interests, intergroup competition can be an engine for cultural evolution.[11] In democratic societies, the act of voting signals the individual's participation in the collective. If citizens adhered to orthodox rationality, they would abstain because their own vote cannot measurably contribute to the victory of their preferred party. However, citizens who reason by the evidential calculus will project their own inclination to vote or to abstain more strongly to their ingroup (i.e., supporters of the same party) than to outgroups (i.e., supporters of other parties). When they decide to vote because this decision signals to them that other ingroup members will probably vote too, the party with the largest number of projectors will prevail. At the same time, the collective public good of a high voter turnout is provided (Acevedo & Krueger, 2004; Quattrone & Tversky, 1984). In other words, in a nested social dilemma such as voting, one would not want people to return to narrowly defined self-interest.

CONCLUSION

As the story of social projection unfolds, theoretical accounts of its basic mental mechanisms become more textured, and the benefits and liabilities of this judgemental heuristic become more evident. As of this writing, social projection appears to be a social-perceptual default mode of thinking that can be engaged with little effort and outside of awareness. At the same time, there appear to be resource-consuming operations that humans can draw on to preserve a sense of uniqueness, and thus personal identity.

The best-documented benefit arising from social projection is the improved accuracy of social perception, and the benefit increases in as much as direct information about others is lacking. Other benefits of projection include increases in attitude certainty (Holtz, 2003), attraction to significant others, and satisfaction with interpersonal relationships (Murray,

[11]Even biological evolution may be stimulated by the mixing of gene pools of groups that would, without conflict or conquest, remain isolated and inbred. Game theorist and Nobel Laureate of economics Robert Aumann advised researchers to avoid moral confusion by not entertaining the moral dimension of conflict in the first place. With respect to war, he suggested "*Don't* try to cure it. Just try to *understand* it" (Aumann, 2006, p. 17075, emphasis in the original).

Holmes, & Griffin, 1996). Problems arise when people fail to realise that their self-referent information is privileged, and thus not projectible. Non-regressive prediction in affective forecasting (Wilson & Gilbert, 2003), projection of unique self-attributes (Birch, 2004), and projection across situational boundaries (van Boven & Lowenstein, 2005) are currently active areas of research.

The model presented here treats social projection as a special case of inductive reasoning. The accuracy of projective predictions can be assessed when criterion measures are available; that is, when the attributes of others are known to the investigators. The philosophy of mind is concerned about an inference problem that is superficially similar to social prediction. This is "the problem of other minds". Augustine, and later John Stuart Mill, made the argument from analogue, according to which people use the phenomenal experience of their own minds as *prima facie* evidence for the idea that other people have minds too (see Malcolm, 1962, for a famous critique). Their own consciousness being, by definition, readily accessible, people are thought to simulate other minds by guessing that they are much like their own. Nichols (in press) has recently noted the relevance of social projection for simulation theory. However, the idea that inferences about the existence of other minds are accurate assumes that other minds do indeed exist. The question is how philosophers could know this without having gone themselves through a round of simulations first. However intuitively compelling this line of reasoning may seem, it cannot be logical.

Social dilemmas are of particular theoretical interest because they break the default operation of social projection. In social dilemmas, people cannot recruit existing self-referent knowledge and project it onto others. Instead, they need to make choices between behavioural strategies that are associated with different interdependent outcomes. Because a person's response may not be settled before the person has inspected the payoff structure, social projection becomes a tool for strategic reasoning. Depending on *what they might do*, people can construct different projections, and thus anticipate different outcomes. If they use their own projective predictions as information relevant for the assessment of expected values, they can overcome the free-rider problem within the group. Now social projection is no longer fully automatic, and it is no longer fully independent of conformity. Instead, evidential decision theory suggests a loop from social projection (i.e., predictions based on one's own presumed cooperation or defection) to conformity (i.e., do that which the majority is thought to do and that which yields the higher personal payoff), and back to projection (i.e., project the resulting choice to others).

The analysis of evidential decision making in social dilemmas has revealed that simple value judgements about behaviour must be suspended. Cooperation can be good or bad depending on whether it is framed in the

context of the individual's perspective, the group's perspective, or the collective's perspective. In the intergroup context, social projection also loses its appeal as a fast, frugal, and beneficial heuristic. If unchecked, differential social projection in nested interpersonal–intergroup dilemmas can be a stimulant of collective disaster. There is no longer a simple mandate to make projection or cooperation larger or smaller. Judgements about what kinds of psychological or behavioural changes are desirable require new assessments that are sensitive to the social contexts in which thought and behaviour occur.

REFERENCES

Acevedo, M., & Krueger, J. I. (2004). Two egocentric sources of the decision to vote: The voter's illusion and the belief in personal relevance. *Political Psychology*, *25*, 115–134.

Acevedo, M., & Krueger, J. I. (2005). Evidential reasoning in the prisoner's dilemma game. *American Journal of Psychology*, *118*, 431–457.

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. D. Alicke, D. Dunning, & J. I. Krueger (Eds.), *The self in social judgement* (pp. 85–106). New York: Psychology Press.

Alicke, M. D., & Largo, E. (1995). The role of the self in the false consensus effect. *Journal of Experimental Social Psychology*, *31*, 28–47.

Allport, F. H. (1924). The group fallacy in relation to social science. *Journal of Abnormal and social Psychology*, *19*, 60–73.

Ames, D. R. (2004a). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology*, *87*, 573–585.

Ames, D. R. (2004b). Inside the mind reader's tool kit: Projection and stereotyping in mental state inference. *Journal of Personality and Social Psychology*, *87*, 340–353.

Attneave, F. (1968). Triangles as ambiguous figures. *American Journal of Psychology*, *81*, 447–453.

Aumann, R. J. (2006). War and peace. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 17075–17078.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Birch, S. A. J. (2004). Current knowledge is a curse: Children's and adults' reasoning about mental states. *Current Directions in Psychological Science*, *14*, 25–29.

Bornstein, G. (2003). Intergroup conflict: Individual, group, and collective interests. *Personality and Social Psychology Review*, *7*, 129–145.

Brams, S. J. (1975). Newcomb's problem and prisoner's dilemma. *Journal of Conflict Resolution*, *19*, 596–612.

Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*, 429–444.

Brewer, M. B. (in press). Depersonalised trust and ingroup cooperation. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn M. Dawes* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology*, *66*, 268–275.

Cadinu, M. R., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal group setting. *Journal of Personality and Social Psychology*, *70*, 661–677.

Chen, F. F., & Kenrick, D. T. (2002). Repulsion or attraction? Group membership and assumed similarity. *Journal of Personality and Social Psychology*, *83*, 111–125.

Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity, and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. II, pp. 151–192). Boston, MA: McGraw-Hill.

Clement, R. W., & Krueger, J. (2000). The primacy of self-referent information in perceptions of social consensus. *British Journal of Social Psychology*, *39*, 279–299.

Clement, R. W., & Krueger, J. (2002). Social categorisation moderates social projection. *Journal of Experimental Social Psychology*, *38*, 219–231.

Coombs, C. (1973). A reparameterization of the Prisoner's Dilemma Game. *Behavioural Science*, *18*, 424–428.

Coser, L. A. (1956). *The functions of social conflict*. Glencoe, IL: Free Press.

Crisp, R. J., & Hewstone, M. (2000). Crossed categorisation and intergroup bias: The moderating roles of intergroup and affective context. *Journal of Experimental Social Psychology*, *36*, 357–383.

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorization. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 39). Orlando, FL: Academic Press.

Dawes, R. M. (1989). Statistical criteria for establishing a truly false consensus effect. *Journal of Experimental Social Psychology*, *25*, 1–17.

Dawes, R. M. (1990). The potential nonfalsity of the false consensus effect. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 179–199). Chicago, IL: The University of Chicago Press.

Dawes, R. M., McTavish, J., & Shaklee, H. (1977). Behaviour, communication, and assumptions about other people's behaviour in a commons dilemma situation. *Journal of Personality and Social Psychology*, *35*, 1–11.

Dawes, R. M., Orbell, J. M., Simmons, R. T., & van de Kragt, A. J. C. (1986). Organising groups for collective action. *American Political Review*, *80*, 1171–1185.

Deutsch, M. (1960). The effect of motivational orientation upon trust and suspicion. *Human Relations*, *13*, 123–139.

DiDonato, T. E., & Krueger, J. I. (2007). *Intergroup differentiation and ingroup favouritism: Evidence for the sufficiency of egocentric processes*. Manuscript in preparation.

Eells, E. (1991). *Probabilistic causality*. Cambridge, UK: Cambridge University Press.

Epley, N., Keysar, B., & van Boven. L. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, *87*, 327–339.

Gaertner, L., & Insko, C. A. (2000). Intergroup discrimination in the minimal group paradigm: Categorisation, reciprocation, or fear? *Journal of Personality and Social Psychology*, *79*, 77–94.

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (2001). Reducing intergroup bias: The benefits of recategorisation. In M. A. Hogg & D. Abrams (Eds.), *Intergroup relations: Essential readings* (pp. 356–369). Philadelphia: Psychology Press.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgement: An egocentric bias in the estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, *78*, 211–222.

Govorun, O., Fuegen, K., & Payne, B. K. (2006). Stereotypes focus defensive projection. *Personality and Social Psychology Bulletin*, *32*, 781–793.

Gould, R. V. (1999). Collective violence and group solidarity: Evidence from feuding society. *American Sociological Review*, *64*, 356–380.

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

Gramzow, R. H., Gaertner, L., & Sedikides, C. (2001). Memory of in-group and out-group information in a minimal group context: The self as an informational base. *Journal of Personality and Social Psychology*, *80*, 188–205.

Holtz, R. (2003). Intragroup and intergroup projection can increase opinion uncertainty: Is there classism in college? *Journal of Applied Social Psychology*, *33*, 1922–1944.

Isnko, C. A., Kirchner, J. L., Pinter, B., Efaw, J., & Wildschut, T. (2005). Interindividual–intergroup discontinuity as a function of trust and categorisation: The paradox of expected cooperation. *Journal of Personality and Social Psychology*, *88*, 365–385.

Jeffrey, R. (1964/1983). *The logic of decision* (2nd ed.). Chicago, IL: University of Chicago Press.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge, UK: Cambridge University Press.

Kahan, J. P. (1974). Rationality, the prisoner's dilemma, and population. *Journal of Social Issues*, *30*, 189–210.

Kawada, C. L. K., Oettingen, G., & Gollwitzer, P. M. (2004). The projection of implicit and explicit goals. *Journal of Personality and Social Psychology*, *86*, 545–559.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*, 25–41.

Krueger, J. (1992). On the overestimation of between-group differences. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 3, pp. 31–56). Chichester, UK: Wiley & Sons.

Krueger, J. (1996). Personal beliefs and cultural stereotypes about racial characteristics. *Journal of Personality and Social Psychology*, *71*, 536–548.

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. (2000). 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. I., & Acevedo, M. (2005). Social projection and the psychology of choice. In M. D. Alicke, D. Dunning, & J. I. Krueger (Eds.), *The self in social perception* (pp. 17–41). New York: Psychology Press.

Krueger, J. I., & Acevedo, M. (in press). Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning. *American Journal of Psychology*.

Krueger, J. I., Acevedo, M., & Robbins, J. M. (2005). Self as sample. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 353–377). New York: Cambridge University Press.

Krueger, J., & Clement, R. W. (1994). The truly false consensus effect: An ineradicable and egocentric bias in social perception. *Journal of Personality and Social Psychology*, *67*, 596–610.

Krueger, J., & Clement, R. W. (1996). Inferring category characteristics from sample characteristics: Inductive reasoning and social projection. *Journal of Experimental Psychology: General*, *125*, 52–68.

Krueger, J. I., & DiDonato, T. E. (2007). *Reputation and last-minute intrigue in the prisoner's dilemma*. Manuscript in preparation.

Krueger, J., & Rothbart, M. (1988). Use of categorical and individuating information in making inferences about personality. *Journal of Personality and Social Psychology*, *55*, 187–195.

Krueger, J., & Stanke, D. (2001). The role of self-referent and other-referent knowledge in perceptions of group characteristics. *Personality and Social Psychology Bulletin*, *27*, 878–888.

Krueger, J., & Zeiger, J. S. (1993). Social categorisation and the truly false consensus effect. *Journal of Personality and Social Psychology*, *65*, 670–680.

Kunda, Z., & Sherman-Williams, B. (1993). Stereotypes and the construal of individuating information. *Personality and Social Psychology Bulletin*, *19*, 90–99.

Levi, I. (2000). Review of "The foundations of causal decision theory" by J. M. Joyce. *Journal of Philosophy*, *97*, 387–402.

Lodewijkx, H. F. M., Rabbie J. M., & Visser, L. (in press). "Better to be safe than sorry": Extinguishing the individual–group discontinuity effect in competition by cautious reciprocation. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology*. Hove, UK: Psychology Press.

Malcolm, N. (1962). Knowledge of other minds. In V. C. Chappell (Ed.), *The philosophy of mind* (pp. 151–159). Englewood Cliffs, NJ: Prentice-Hall.

Messé, L. A., & Sivacek, J. M. (1979). Predictions of others' responses in a mixed-motive game: Self-justification or false consensus? *Journal of Personality and Social Psychology*, *37*, 602–607.

Moors, A., & de Houwer, J. (2007). What is automaticity? An analysis of its component features and their interrelations. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 11–50). New York: Psychology Press.

Mullen, B., & Hu, L. (1989). Perceptions of ingroup and outgroup variability: A meta-analytic integration. *Basic and Applied Social Psychology*, *10*, 233–252.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality & Social Psychology*, *70*, 79–98.

Newman, L., Duff, S., Kimberly, J., & Baumeister, R. F. (1997). A new look at defensive projection: Thought suppression, accessibility, and biased person perception. *Journal of Personality and Social Psychology*, *72*, 980–1001.

Nichols, S. (in press). Mindreading and the philosophy of mind. In J. Prinz (Ed.), *The Oxford handbook on philosophy of psychology*. New York: Oxford University Press.

Otten, S. (2002). "Me" and "us" or "us" and "them"? The self as a heuristic for defining minimal ingroups. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 13, pp. 1–34). Hove, UK: Psychology Press.

Otten, S., & Epstude, K. (2006). Overlapping mental representations of self, ingroup, and outgroup: Unraveling self-stereotyping and self-anchoring. *Personality and Social Psychology Bulletin*, *32*, 957–969.

Otten, S., & Wentura, D. (2001). Self-anchoring and in-group favouritism: An individual profiles analysis. *Journal of Experimental Social Psychology*, *37*, 525–532.

Poundstone, W. (1992). *Prisoner's dilemma*. New York: Doubleday.

Pronin, E., Wegner, D. M., McCarthy, K., & Rodriguez, S. (2006). Everyday magical powers: The role of apparent mental causation in the overestimation of personal influence. *Journal of Personality and Social Psychology*, *91*, 218–231.

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, *28*, 363–392.

Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, *46*, 237–248.

Rabbie, J. M., Schot, J. C., & Visser, L. (1989). Social identity theory: A conceptual and empirical critique from the perspective of a behavioural interaction model. *European Journal of Social Psychology*, *19*, 171–202.

Rapoport, A. (1967). A note on the index of cooperation for Prisoner's Dilemma. *Journal of Conflict Resolution*, *11*, 101–103.

Reichenbach, H. (1956). *The direction of time*. Berkeley, CA: University of California Press.

Riketta, M. (2006). Projection of self-attributes to outgroups. In A. P. Prescott (Ed.), *The concept of self in psychology* (pp. 215–241). Hauppauge, NY: Nova Science Publishers.

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.

Ross, L., Greene, D., & House, P. (1977). The false consensus effect: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*, 279–301.

Rubin, M., & Badea, C. (2007). Why do people perceive ingroup homogeneity on ingroup traits and outgroup homogeneity on outgroup traits? *Personality and Social Psychology Bulletin*, *33*, 31–42.

Russell, B. (1932). On the notion of cause. In B. Russell (Ed.), *Mysticism and logic, and other essays* (pp. 180–208). London: Allen & Unwin.

Russell, B. (1959). *Common sense and nuclear warfare*. New York: Simon & Schuster.

Ryan, C. S., & Bogart, L. M. (2001). Longitudinal changes in the accuracy of new group members' in-group and out-group stereotypes. *Journal of Experimental Social Psychology*, *37*, 118–133.

Sally, D. (1995). Conversation and cooperation in social dilemmas. *Rationality and Society*, *7*, 58–92.

Schelling, T. C. (1960). *The strategy of conflict*. New York: Oxford University Press.

Schimel, J., Greenberg, J., & Martens, A. (2003). Evidence that projection of a feared trait can serve a defensive function. *Personality and Social Psychology Bulletin*, *29*, 969–979.

Schlenker, B. R., & Bonoma, T. V. (1978). Fun and games: The validity of games for the study of social conflict. *Journal of Conflict Resolution*, *22*, 7–38.

Shafir, E., & Tversky, A. (1992). Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology*, *24*, 449–474.

Stouten, J., De Cremer, D., & van Dijk, E. (2006). Violating equality in social dilemmas: Emotional and retributive reactions as a function of trust, attribution, and honesty. *Personality and Social Psychology Bulletin*, *32*, 894–906.

Stroebe, K., Lodewijkx, H. F. M., & Spears, R. (2005). Do unto others as they do unto you: Reciprocity and social identification as determinants of ingroup favouritism. *Personality and Social Psychology Bulletin*, *31*, 831–845.

Tajfel, H. (1959). Quantitative judgement in social perception. *British Journal of Psychology*, *50*, 16–29.

Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, *25*, 79–97.

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.

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

Van Boven, L., & Loewenstein, G. (2005). Cross-situational projection. In M. D. Alicke, D. Dunning, & J. I. Krueger (Eds.), *The self in social judgement* (pp. 43–64). New York: Psychology Press.

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, *77*, 337–349.

Van Lange, P. A. M. (2000). Beyond self-interest: A set of propositions relevant to interpersonal orientations. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 11, pp. 279–331). Chichester, UK: Wiley.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345–411). San Diego, CA: Academic Press.

Wildschut, T., & Insko, C. A. (2006). A paradox of individual and group morality: Social philosophy and empirical philosophy. In P. A. M. van Lange (Ed.), *Bridging social psychology: Benefits of transdisciplinary approaches* (pp. 377–384). Mahwah, NJ: Erlbaum.

Wildschut, T., Pinter, B., Vevea, J. L., Insko, C. A., & Schopler, J. (2003). Beyond the group mind: A quantitative review of the interindividual–intergroup discontinuity effect. *Psychological Bulletin*, *129*, 698–722.

Yamagishi, T., & Kiyonari, T. (2000). The group as the container of generalised reciprocity. *Social Psychology Quarterly*, *63*, 116–132.