Feeling "Holier Than Thou": Are Self-Serving Assessments Produced by Errors in Self- or Social Prediction?

Nicholas Epley and David Dunning
Cornell University

People typically believe they are more likely to engage in selfless, kind, and generous behaviors than their peers, a result that is both logically and statistically suspect. However, this oft-documented tendency presents an important ambiguity. Do people feel "holier than thou" because they harbor overly cynical views of their peers (but accurate impressions of themselves) or overly charitable views of themselves (and accurate impressions of their peers)? Four studies suggested it was the latter. Participants consistently overestimated the likelihood that they would act in generous or selfless ways, whereas their predictions of others were considerably more accurate. Two final studies suggest this divergence in accuracy arises, in part, because people are unwilling to consult population base rates when predicting their own behavior but use this diagnostic information more readily when predicting others'.

> God, I thank thee that I am not like other men, extortioners, unjust, adulterers, or even like this tax collector.
>
> Luke 18:11, *Revised Standard Version*

On February 7, 1998, at the dawn of what would become only the second presidential impeachment in American history, CBS News conducted a telephone poll asking Americans how interested they were in the steamy details of Bill Clinton's sex life. The results suggested they were not very interested. A scant 7% reported being "fascinated," and 50% reported being "completely disinterested." However, when asked to evaluate the interest of other Americans, 25% thought others were "fascinated" and only 18% thought others were "completely disinterested." Clearly, most people felt their own ambivalence was unique and that the hysteria surrounding the Clinton affair was primarily driven by the vulgarity of others. When presented with this blatant discrepancy, one polling expert noted, "It's the great contradiction: the average person believes he is a better person than the average person" (Berke, 1998, p. 4:1).

It is a contradiction, indeed, but certainly not one that is new to social psychologists. Researchers have repeatedly demonstrated that people on average tend to think they are more charitable, cooperative, considerate, fair, kind, loyal, and sincere than the typical person but less belligerent, deceitful, gullible, lazy, impolite, mean, and unethical—just to name a few (Alicke, 1985; Allison, Messick, & Goethals, 1989; Dunning, Meyerowitz, & Holzberg, 1989; Goethals, Messick, & Allison, 1991).

These self-serving assessments are quite general, appearing not only in beliefs about abstract traits and abilities but also in predictions of specific behavior. For example, people generally think they are more likely than their peers to rebel in the Milgram obedience studies, cooperate in a prisoner's dilemma game, distribute collective funds equitably, and give up their seat on a crowded bus to a pregnant woman (Bierbrauer, 1976; Goethals et al., 1991). In addition, people tend to believe they will resolve moral dilemmas by selecting the saintly course of action but that others will behave more selfishly (Allison et al., 1989). Although people's general tendency to think they were less interested than others in the president's sex life is something of a contradiction, it is definitely not an anomaly. In fact, people seem to chronically feel "holier than thou."

The most extensive demonstration of this bias in predictions of behavior comes from work on the uniqueness bias, the tendency for people to underestimate the commonness of their own desirable behaviors (for reviews, see Goethals, 1986; Goethals et al., 1991). In one representative study (Goethals, 1986), students were asked whether they would be willing to donate blood at a local charity event and also to estimate the percentage of their peers who would donate blood if given the chance. Overall, 60% said they would be

willing but that only 39% of their peers would feel the same. In a similar study (Goethals, 1986), students in an introductory psychology class were informed about a standard prisoners' dilemma game and asked whether they would behave cooperatively or competitively. Overall, 63% indicated they would behave cooperatively, but these cooperative students thought only 35% of their peers would behave likewise.

Although demonstrations of self-serving predictions of behavior are numerous and robust, they leave open an important ambiguity. When people make logically impossible predictions about themselves in relation to others, which prediction is in error? Comparisons of oneself with others involves two components. First, people must evaluate their own traits, dispositions, and likely behavior. Second, they must evaluate the traits, dispositions, and likely behavior of others. An error in either component could lead to flattering "holier than thou" assessments. Perhaps people really do have privileged access to the workings of their own minds (Lewis, 1952) and evaluate themselves relatively accurately but are overly cynical and harsh in their expectations of their peers. Or, perhaps people more dispassionately and accurately assess the dispositions of others and hold overly charitable views about themselves.

The research reported here was primarily aimed at determining which error, if not both, produces the flattering self-assessments that occur when people compare themselves with others. To be sure, Goethals and colleagues stated that their respondents felt uniquely moral because they underestimated "the proportion of people who can or will perform socially desirable actions" (Goethals et al., 1991, p. 149), but two issues cloud this conclusion. First, as seen in the examples discussed above, Goethals and colleagues (1991) tended to compare how participants claimed they would behave in a given situation with their estimates of how their peers would actually behave in that same situation. This procedure left unanswered whether claims about one's own behavior would prove accurate if those hypothetical situations became reality. Second, in the studies in which actual behavior was observed, that behavior immediately followed a session in which people made their predictions (Goethals, 1986). Actual behavior in these cases may have been contaminated by the well-documented finding that predicting one's behavior tends to alter that behavior significantly. Once a person makes a prediction about a course of action, they tend to stick with it more resolutely than if no prediction had been made (Greenwald, Carnot, Beach, & Young, 1987; Sherman, 1980). An additional goal of the present research was to determine precisely the mechanisms that may systematically produce the accuracy and error we uncovered in the initial studies.

To address these goals, we asked participants to predict how they and their peers would behave in situations with moral or altruistic overtones. These included decisions regarding whether to buy flowers to benefit a national charity (Study 1), to cooperate in a prisoners' dilemma game (Study 2), to donate money to a nonprofit organization (Study 3), and to assign oneself instead of a partner to an unpleasant experimental task (Study 4). These hypothetical predictions were compared with the behavior of participants who actually experienced the dilemma to determine whether self-serving comparisons were produced by mistaken predictions of one's own or others' behavior.

We focused on behavior in the moral domain, in part because predictions involving morality, kindness, generosity, and fairness are of obvious importance in social life. Also, nowhere is the tendency to make flattering self-assessments more robust than in domains charged with moral or altruistic overtones. Whereas people may be somewhat reluctant to say they are smarter than their peers, they have no difficulty noting that they are more generous, fair, ethical, or moral (Allison et al., 1989; van Lange, 1991; van Lange & Sedikides, 1998). Indeed, despite all of the talk in the United States about a "moral majority," the majority of Americans consider themselves to be elite members of a "moral minority."

Study 1: Daffodil Days

As an initial investigation into the accuracy of self- and social predictions, we asked students at Cornell University to predict their behavior during an annual campus charity event. On one extended weekend every spring, fraternities and sororities at Cornell University sell daffodils to benefit the American Cancer Society. This event typically lasts 4 days, from Thursday through Sunday. During these "Daffodil Days," fraternity and sorority members swarm the campus, soliciting sales and donations on nearly every campus intersection as well as several dozen points throughout the community.

Roughly 5 weeks before the 1999 version of Daffodil Days, 251 students enrolled in an upper-level psychology course were presented a questionnaire reminding them of the upcoming charity event, describing its purpose, and detailing the cost of individual daffodils. Students then predicted whether or not they would buy at least one daffodil during the 4-day event and, if so, how many they would buy. Students also predicted the percentage of their peers enrolled in the same psychology course who would buy at least one and how many, on average, each of their peers would purchase (including those students who would not buy any). These questions were attached to the end of a class examination. Students were awarded extra credit toward their course grade for their participation.

To assess the accuracy of these predictions, these same students were presented with another questionnaire on a second course examination given 3 days after the conclusion of Daffodil Days. This questionnaire simply asked each student to indicate whether or not he or she had purchased at least one daffodil during the charity drive and, if so, how many were purchased. Once again students were give extra credit for their participation, and only those who completed both prediction and actual behavior questionnaires (79% of the students enrolled in the course) were included in the following analyses.

As expected, students felt "holier than thou" before Daffodil Days. A total of 83% predicted they would buy at least one flower but that only 56% of their peers, on average, would do likewise, paired $t(250) = 11.44$, $p < .001$. Furthermore, students predicted they would buy two flowers, on average, whereas their peers would buy only 1.6, paired $t(238) = 3.86$, $p < .001$.[1]

---

[1] Degrees of freedom vary for the tests involving the number of daffodils purchased because some students failed to provide responses on this item or provided unreasonably high estimates for their peers. Ten students fell into this latter category (4% of total sample) and indicated that each of their peers would buy anywhere from 32 to 501 daffodils. These responses suggested they misinterpreted our question as the total number their peers would purchase collectively rather than the average number each student would purchase individually. Aside from these participants, the highest number predicted for peer purchases was 6.

Despite the fact that 83% of students predicted they would buy at least one daffodil, only 43% actually did so. In addition, despite predicting that they would purchase, on average, two flowers each, these students instead purchased 1.2. A comparison between predicted and actual behavior indicated these students substantially overestimated the likelihood that they would buy at least one flower as well the number of flowers they would purchase, paired $ts(250) = 11.79$ and 5.90, respectively, $ps < .001$. Predictions about others, however, were considerably more accurate. Students only slightly, but still significantly, overestimated the percentage who would buy at least one daffodil as well as the number they would purchase, paired $ts(250$ and $236) = 4.09$ and 2.20, respectively, $ps < .05$.

These data provide initial evidence that self-serving assessments primarily reflect errors in self- rather than social prediction. Students significantly overestimated both the likelihood and number of daffodils they would actually buy but were relatively more accurate when predicting their peers. Furthermore, the completely within-participant design of this study made it easy to identify both errors and accuracy in prediction, as those who predicted their behavior were also those who actually experienced the situation.

This design, however, is not without concerns. On the one hand, errors in self-prediction could have been inflated by mistaken construals of the situation surrounding Daffodil Days (Griffin, Dunning, & Ross, 1990; Griffin & Ross, 1991). When making their predictions, students may have brought to mind a situation that differed significantly in its objective details from the one actually encountered. Students may have imagined, for example, being calmly greeted by a charming fraternity brother at the precise moment when they happened to be carrying some "extra" cash, when in fact they were more likely to be assertively confronted by an overbearing college student just before converting their last borrowed dollar into a vending machine lunch. Although mistaken construals are unlikely to explain why students felt "holier than thou," they may have inflated the magnitude of prediction errors for both oneself and others.

On the other hand, it is also possible that the within-participants design of this study may have actually diminished the magnitude of prediction errors. After all, the very act of predicting one's own behavior may lead people to behave in a manner that confirms that prediction (Greenwald et al., 1987; Sherman, 1980). Although the potential for contamination seemed minimal in this study as predictions were made more than a month in advance of the behavior, it is possible that student's predictions influenced their subsequent behavior and that this study may have underestimated the magnitude of prediction errors.

Study 2: A Saint's Dilemma

To avoid the possibility of contamination and reduce the potential for construal differences, we adopted a between-participants methodology in the remaining studies. In this design, participants were randomly assigned to conditions in which they either predicted their own behavior and that of their peers or experienced the actual situation without making prior predictions. Because the prediction and actual behavior conditions involved different participants, there was no chance for predictions to influence behavior or vice versa. Also, because the actual situation was described in great detail, it was not possible for predictors to imagine a situation

that differed, at least in its objective features, from the situation actually encountered.

In the first experiment with this design, we adapted a study by Goethals (1986) that used a time-honored experimental paradigm, the prisoners' dilemma game. After being shown the standard choices and payoff structure of the game, roughly half of the participants predicted how they and their peers would respond to the dilemma (i.e., whether they would choose to cooperate or defect). The other half, without making predictions, confronted the actual dilemma and made the choice. Given our findings from Study 1, we anticipated that participants would once again feel "holier than thou," predicting they would behave more cooperatively than their peers, and that their saintly self-predictions would be less accurate than their more cynical predictions of others when compared with the group who actually experienced the prisoner's dilemma.

Method

Participants

A total of 97 undergraduates at Cornell University participated for extra credit in their psychology and human development courses.

Procedure

Participants were run in groups of 4 to 12 and randomly separated into prediction and actual behavior conditions on arrival to the laboratory. Those in the actual behavior condition ($n = 46$) were told they would be playing a game that would allow them the opportunity to earn money. They were informed that this game was to be played in pairs and that they would be randomly assigned to another participant in the experiment whose identity would remain unknown. The outcomes of the game, they were told, would depend on their own and their partner's behavior. The prediction group ($n = 51$) was given identical instructions except they were told that the game was purely hypothetical and that they were to anticipate, as best they could, how they would behave if they actually played the game. The experimenter stressed to all participants that their responses would remain completely confidential.

Following this introduction, all participants were separated into private cubicles and given a questionnaire containing a standard two-player prisoners' dilemma game.[2] All participants were told that both they and their partner would have the option to cooperate or defect. As in all prisoners' dilemma games, the payoffs were structured such that defection yielded the higher payoff for each participant regardless of his or her partner's decision but that both fared better if each cooperated. The financial outcomes were carefully described, and a figure of the payoffs was included (see Figure 1).

After reading these instructions, all participants indicated whether they would choose to cooperate or defect on the first round and how many of their peers (Cornell undergraduates) in a random sample of 100 would choose to cooperate and defect. Whether participants first made predictions

---

[2] As an attempt to manipulate the salience of moral and self-interest concerns, the description of the game was either titled the "Interpersonal Trust Game" or the "Interpersonal Economics Game." Participants read that the experiment was investigating how people make decisions either in a trustworthy or untrustworthy manner (trust description) or in an economically rational or irrational manner (economics description). Quite unexpectedly, given previous research using a somewhat similar procedure (Ross & Samuels, 1993, as cited in Ross & Ward, 1995), this manipulation did not influence either hypothetical predictions or actual behavior. As a result, all analyses are collapsed across this factor.

| | Partner's Decision | |
Your Decision:	Cooperate	Defect
Cooperate	You: $7.50 Partner: $7.50	You: $0 Partner: $10
Defect	You: $10 Partner: $0	You: $2.50 Partner: $2.50

Figure 1. Payoff matrix in prisoners' dilemma game (Study 2).

about themselves or others was completely counterbalanced in this and in all subsequent studies. This manipulation did not influence any of the results reported in this article and is not mentioned further.

Results and Discussion

As expected, participants predicted they would be more likely to cooperate than their peers. In fact, the vast majority predicted they would cooperate (84%) but thought only a moderate majority of their peers would do likewise ($M = 64\%$), paired $t(50) = 4.68, p < .0001$. In actuality only 61% chose to cooperate, a figure that did not exceed the actual rate predicted for others, $t(50) = 1.13, p > .2$, but was a far cry from the high rate predicted for themselves, $\chi^2(1, N = 97) = 6.78, p < .01$.[3]

These findings are consistent with those from Study 1, suggesting that self-serving behavioral assessments are produced primarily by errors in self- rather than social prediction. In fact, the same errors in self-prediction occurred even though all participants who made predictions knew the objective circumstances involved in the actual situation. All participants knew the exact payoff structure of the game, thus diminishing differential construal of the objective details of the situation (Griffin & Ross, 1991) as an explanation for prediction errors.

Study 3: Can You Spare a Dime?

The fact that similar results were obtained across two different domains and in different paradigms suggests that this pattern of self-distortions may be more the rule than the exception. Study 3 was designed, in part, to investigate the generalizability of this effect still further by confronting participants with an opportunity to donate any or all of their payment for participating in an experiment to charity.

Study 3 was also designed to investigate one potential misunderstanding that might lead to errors in prediction. In particular, people sometimes have difficulty accurately imagining how their behavior is influenced by emotional or visceral forces that can be felt only when in the actual situation (Loewenstein, 1996; Loewenstein & Adler, 1995; Van Boven, Dunning, & Loewenstein, 2000). As a result, participants may have made prediction errors because they could not anticipate how they would feel if placed in the actual situation. Students in the Daffodil Days experiment, for example, may have been unable to appreciate how

much it would "hurt" to part with money for a mere flower. Although misunderstanding the impact of visceral factors may not account for the difference in predictions for oneself and others, it might help to explain why people have difficulty predicting their own behavior.

We investigated this issue in Study 3 by manipulating the extent to which participants were psychologically remote from the situation under consideration. In the "distant" condition, participants predicted how much of their participation payment they would donate to a charity at the beginning of the experimental session after reading a description of the situation experienced by those in an actual behavior condition. In the "close" condition, participants made their predictions after completing 20 min of unrelated questionnaires and receiving their fee—the exact moment that participants in the actual behavior condition confronted a genuine choice. We presumed that participants in this close condition, introduced to the identical situation facing participants making actual choices, would have more insight into the visceral factors that might influence that choice and might therefore be more accurate in their predictions.

Method

Participants

Thirty-eight high-school and college students attending a summer session at Cornell University were paid $5 for their participation.

Procedure

Each participant was randomly assigned to one of three conditions. In the actual behavior condition ($n = 13$), participants first completed a

---

[3] Participants in the hypothetical condition also indicated the likelihood that they would cooperate or defect as well as the likelihood that the typical Cornell student would cooperate or defect. These ratings were made on 11-point scales ranging from *100% likely to defect* to *100% likely to cooperate*. The midpoint was labeled *50/50: don't know/uncertain*. Once again, participants felt holier than thou, thinking they were more likely to cooperate ($M = 75\%$) than their peers ($M = 63\%$), paired $t(50) = 3.69, p < .001$. Furthermore, one-group t tests revealed that participants significantly overestimated the likelihood they would cooperate compared with the actual likelihood of cooperation (61%), $t(50) = 3.74, p < .001$, but were very accurate when predicting others, $t(50) < 1$.

20-min packet of questionnaires unrelated to the present research. When finished, participants returned their packet to the experimenter, who handed them an unsealed envelope. This envelope, they were told, contained both their payment for participating in the experiment ($5) as well as a sheet of information they should read. Participants returned to their cubicles to do so.

Inside the envelope was a short request asking participants, if they wished, to donate some of their payment to one of three charities: The Salvation Army, The American Red Cross, or the Society for the Prevention of Cruelty to Animals. Promotional materials from each of the charities were included with the request in case participants were unfamiliar with them. If they wished to donate, participants indicated their favored charity, deposited the donation inside the envelope, and returned it (sealed) to the experimenter. If they did not wish to donate, participants returned the envelope and its contents (sealed) to the experimenter. Because everyone returned an anonymous, sealed envelope, the experimenter was unable to tell who had chosen to donate and who had not—a feature we had emphasized to all participants. The actual amount of money donated by these participants served as the baseline against which predictions were compared.[4]

All procedures were identical for participants in the close condition (n = 12), who made predictions after being paid, except that the donation request was explicitly hypothetical. They were told to indicate how much, if any, they would donate if actually given the opportunity to do so and to make the same prediction for their peers (a random sample of 100 Cornell summer students).

Participants in the distant condition (n = 13), who made predictions before being paid, experienced a somewhat different procedure. These participants first read a detailed description of the experiment actually experienced by participants in the actual and close conditions, as well as a description and copy of the donation request. The same promotional materials on each of the three charities were paper clipped to the description and request. These participants then predicted how much they would be willing to donate (and keep) if they had actually participated in this experiment and the amount that their peers (a random sample of 100 Cornell summer students) would choose to donate (and keep).

Results and Discussion

Data Transformations

Participants in this experiment could predict or actually donate anywhere from $0 to $5. Unfortunately for the charities involved, most participants used the miserly end of this scale. There were, however, 5 participants who predicted they would donate all of their payment and 2 who actually did, creating a positively skewed distribution. To restore normality (particularly in the actual behavior condition), all predicted and actual donations were subjected to square-root transformations. All analyses were performed on these transformed values, but the untransformed means are presented in the text and figures to ease interpretation.

Predicted Versus Actual Donations

As expected, participants in the two prediction conditions once again felt "holier than thou," this time anticipating they would donate more money than their peers (Ms = $2.44 and $1.83, respectively). A 2 (predicting: self or peers) × 2 (condition: distant or close) mixed-model analysis of variance (ANOVA) indicated that this difference was significant, $F(1, 23)$ = 4.19, p = .05. Neither the main effect for condition nor the interaction was significant, $Fs(1, 23)$ = 1.57 and 0, respectively, ps > .2. Although participants in the close condition did think they would donate less

Table 1

Predicted Versus Actual Amount of Donation for Self and Others in the Close and Distant Conditions

Condition	Donation amount	
	Predicted for self	Predicted for others
Distant	$2.75	$2.03
Close	$2.15	$1.69
Average	$2.44	$1.83
Actual behavior	$1.53	

Note. Participants could predict to donate or actually donate anywhere from $0 to $5.

than those in the distant condition, they did not do so significantly, $t(23)$ = 1.03, p > .3 (see Table 1). This somewhat unexpected result suggests that the pattern of errors observed in Studies 1 and 2 is fairly robust. All further analyses are collapsed across this factor.

As in the previous studies, participants' charitable self-predictions were less accurate than their predictions for others. On average, participants actually donated $1.53. Planned contrasts comparing these predictions against the actual behavior of those confronted with the dilemma revealed that the participants' self-predictions were significantly higher than the actual donations, $t(35)$ = 2.74, p = .01, but predictions of others were only marginally higher, $t(35)$ = 1.78, p = .09. Although participants generally overestimated the amount that both they and their peers would donate, they did so more when predicting their own behavior.

Participants also anticipated how many of their peers would donate at least some portion of their payment. Every participant in the prediction conditions anticipated he or she would donate something but that only 51% of their peers, on average, would do likewise. In actuality, 62% chose to donate. Participants therefore overestimated the likelihood that they themselves would act charitably, $\chi^2(1, N = 38)$ = 11.08, p < .001, and slightly, albeit significantly, underestimated the likelihood that their peers would donate as well, $t(24)$ = −3.05, p < .01.

As in Studies 1 and 2, self-serving predictions were produced more by errors in self- rather than social prediction. In this experiment, participants considerably overestimated how much they would actually donate but were relatively more accurate when predicting others. Furthermore, manipulating the psychological closeness of participants making predictions did not increase their accuracy or significantly attenuate their tendency to feel "holier than thou."

Study 4: Can You Spare the Time?

Study 4 was designed as a more detailed investigation into the relative accuracy of self- versus social predictions in the moral domain. In the first three studies, accuracy in self- and social predictions was assessed by presenting participants with one moral dilemma. In Study 4, we introduced situational variations into a

---

[4] All money actually donated in this experiment was delivered to each participant's chosen charity.

dilemma and asked participants to anticipate how these variations would influence their own and others' behavior. These variations allowed us to determine not only whether people could more accurately predict others' behavior in moral dilemmas but also whether they were better able to track changes in the actual base rates of others' behavior.

These situational variations involved manipulating the strengths of two motivational forces that were in basic competition in the situations presented to participants in the first three studies. On the one hand, each study aroused what Adam Smith (1759/1976) would call "moral sentiments," thoughts about the well-being of others and an assessment of the "right" or "moral" course of action. On the other hand, each study also aroused concerns about material self-interest, in each case how much money the participant would walk away with.

In Study 4 we manipulated the relative strength of these two concerns to investigate the extent to which participants could correctly anticipate how much moral sentiments and self-interest actually influenced their own and others' behavior. In this study, participants were asked to assign either themselves or a partner to a pleasant or an unpleasant experimental task. Moral sentiments were manipulated by altering the supposed identity of the participant's partner (either a typical college student or a 10-year-old girl), and self-interest was manipulated by altering the time required to complete the unpleasant task (8 min or 30 min). Because self-interest commonly exerts considerable influence over actual behavior (Green & Cowden, 1992), we suspected that those who actually confronted our experimental dilemma would be more strongly influenced by variations in self-interest concerns than by variations in moral sentiments. On the basis of the first three experiments, we expected that participants would track this actual influence more accurately when predicting others' behavior than when predicting their own.

Method

Participants

A total of 61 Cornell undergraduates participated for extra credit in their psychology or human development courses. Of these, 29 predicted their own and others' reactions to the dilemma and 32 actually experienced it.

Procedure in Actual Behavior Condition

All participants were told the experiment investigated the accuracy of "social impressions," in particular how the accuracy of these impressions changes across the lifespan. Participants were further told they had been recruited with a partner, someone they were unlikely to know, who was sitting in another area of the lab but whom they would never actually meet. Participants learned they would first be asked to complete some questionnaires detailing their impressions of themselves. A picture of them would then be given to their partner who would make evaluations based solely on this picture, and they would do likewise with their partner's picture.

As promised, participants then evaluated themselves along 12 bipolar trait dimensions, had their picture taken, and rated their partner on the same dimensions based solely on a Polaroid snapshot. Once finished, the experimenter informed participants that there were two different tasks that would need to be completed in the second part of the experiment, one by the participant and the other by his or her partner. The experimenter explained that because she needed to remain unaware of the person's condition, she was unable to assign them to complete the tasks herself and had instead

randomly selected the participant to do so. The experimenter then handed participants a large envelope containing a written description of both tasks. Participants returned to a private cubicle, read the two descriptions, removed the one they chose to complete, and sealed the envelope with the description of their partner's task inside. Once participants had returned the sealed envelope, the experiment was terminated and the participants fully debriefed.

Manipulating self-interest. Participants were asked to assign either themselves or their partner to one of two tasks, respectively titled Video Clips A and B. The description of Video Clips A was identical in the high and low self-interest conditions. It read,

> In the second part of this experiment, you will be asked to watch several video clips of people engaging in various behaviors. You will then be asked to make a variety of judgments about the actors in these video clips. There are a total of 5 clips that you will be asked to evaluate.
>
> This task is fairly short and will take approximately 5–6 minutes to complete. When finished you will be given a short explanation of the study and will then be free to leave.

The description of the second task, Video Clips B, was identical in all respects but the length of time. In the low self-interest condition, the task was described as "not very long" and that it would take 8–10 min to complete. In the high self-interest condition, the task was described as "fairly long" and that it would take 30 min to complete.[5] We reasoned, quite simply, that people would be more motivated to choose the short task for themselves when the alternative was substantially longer than when it was only trivially longer.

Manipulating moral sentiment. The amount of moral sentiment was manipulated by altering the picture of the participants' partner. In the low moral sentiment condition, participants received a picture of a rather typical male college student. In the high moral sentiment condition, participants received a picture of a 10-year-old girl. We reasoned that participants would feel more guilty assigning the long task to the young girl than to the college student. This reasoning was confirmed by an independent sample of 20 college students, 19 of whom said they would feel more guilty assigning the long task to the young girl (binomial $p < .0001$).

Procedure in the Prediction Condition

Participants who predicted their behavior experienced the same situation, with a few notable exceptions. First, these participants were told that, contrary to the normal sessions, their partner was not actually in the lab with them and as a result they had been randomly paired with a participant from a previous session. Second, because participants knew their partner was not actually in the lab, they were asked to imagine, as best they could, that they had actually been given the task of assigning themselves and their partner to the experimental tasks, and to anticipate their own and others' behavior.

Finally, participants made behavioral predictions in all four cells in the design—High and Low Self-Interest × High and Low Moral Interest—for both themselves and others. After learning about the task assignment, participants received written descriptions of the two tasks along with a large envelope and a response sheet. The response sheet asked participants to indicate which of the two tasks they would assign to themselves and which they would assign to their partner. When finished, participants placed their response sheet inside the envelope and signaled to the experimenter, who was standing around the corner, that they were finished.

After making his or her first prediction, the experimenter mentioned that she would like the participant to imagine that the experiment had been

---

[5] All participants were recruited for a 40-min experiment to ensure that they would theoretically have the time to complete either of the long tasks.

slightly different. Based on a random schedule, the experimenter introduced either a new picture or a new set of tasks and handed the participant another response sheet. Once again, the participant made a prediction and placed the sheet in the envelope. All participants made a total of four predictions about their own behavior, one in each cell of the experimental design.

Participants also made the same predictions in a similar procedure for their peers (a random sample of 100 Cornell University undergraduates). For each cell of the design, participants indicated how many of their peers (out of 100) would select each task. As in the previous studies, whether participants first made predictions about themselves or others was counterbalanced but did not qualify any of the results and is not discussed further.

Results and Discussion

We expected that manipulating the strength of self-interest concerns would influence participants' actual behavior more than manipulating the strength of moral sentiments. As can be seen in Figure 2, this hypothesis was confirmed: Only the self-interest manipulation influenced behavior. Overall, 52% of those actually faced with the dilemma accepted the longer task when self-interest was low (i.e., when the longer task was only trivially so), but only 13% did so when self-interest was high, $\chi^2(1, N = 29) = 5.24, p < .05$. Moral sentiments failed to impact participant's actual behavior in any detectable way, producing neither a main effect nor an interaction on the likelihood of choosing the longer task, both $\chi^2(1, N = 29) < 1$.

Figure 2 also demonstrates, once again, that people's predictions of their own behavior were less accurate than their predictions of others. Across the four situations, participants predicted (on average) that they would volunteer for the long task 57% of the time but that their peers would do so only 40% of the time, $F(1, 28) = 11.56, p < .01$. This admirable self-estimate was far above the 31% of participants in the actual behavior condition who truly chose the long task,

$t(28) = 4.14, p < .001$, revealing that participants once again erroneously thought they would behave more admirably than they actually would. As in Studies 1–3, participants' predictions about their peers were more calibrated, although they still overestimated others' selflessness in this case, $t(28) = 2.26, p < .05$.

Moving beyond the preceding studies, participants in this experiment were also better able to track changes in actual behavior when predicting for others than when predicting for themselves. Planned contrasts within a 2 (self-interest: high or low) × 2 (moral sentiments: high or low) × 2 (target: self or others) repeated-measure design indicated that participants correctly anticipated that others would be strongly influenced by self-interest, $F(1, 28) = 18.03, p < .001$, but erroneously predicted that they themselves would not, $F(1, 28) < 1$. This interaction, however, failed to attain statistical significance, $F(1, 28) < 1$. On the other hand, participants inaccurately predicted that moral sentiments would influence both their own, $F(1, 28) = 27.18, p < .0001$, and others' behavior, $F(1, 28) = 16.91, p < .001$. Although participants generally overestimated the impact of moral sentiments, they did so particularly when predicting their own behavior, interaction $F(1, 28) = 4.93, p < .05$ (see Figure 2). These results indicate that people are not only better able to predict others' behavior in moral dilemmas than their own, but also that they are better able to track changes in the actual base rates of behavior when predicting others' behavior than when predicting their own.

In sum, Study 4 replicated key findings. Participants again predicted that they were more likely to perform a selfless act than their peers, a prediction that again was less accurate for the self than for others. Furthermore, Study 4 demonstrated that participants were generally better able to track how changes in situational features would influence the behavior of others than how those same changes would influence their own behavior. Participants
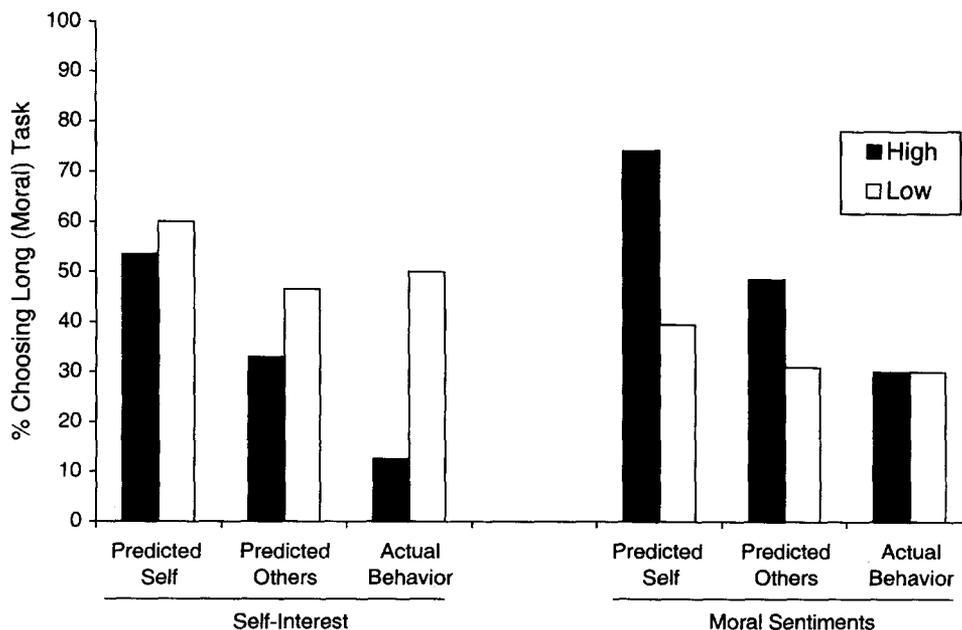

Figure 2. Predicted impact of moral sentiment and self-interest manipulations on oneself and others compared with the actual impact.

tended to predict that their own behavior would be influenced only by the level of moral sentiments inherent in the situation and not by the level of self-interest. Actual behavior, in contrast, was influenced only by participants' self-interest. This pattern of behavior was predicted more accurately in peer predictions, in which participants stated that behavior would be influenced by a mixture of both moral and self-interested concerns. These predictions for others were still overly charitable when compared with actual behavior but, on the whole, tracked actual base rates more accurately than predictions about the self.

Explaining (In)Accurate Predictions of the Self: Internal and External Perspectives in Prediction

Taken together, these studies suggest that people may find it more difficult to accurately predict their own behavior in moral dilemmas than to accurately predict others'. In the first three studies, participants overestimated the likelihood that they would choose the kinder and more generous course of action when faced with altruistic situations by an average of 32% (range = 26%–43%). For others, the average overestimation was a mere 4% (range = 11%–13%). In Study 4, participants erroneously believed that their own behavior would be influenced by the strength of their moral sentiments and not by their self-interest. Participants predicted that their peers would be influenced by both moral and self-interest concerns, a prediction that mapped more closely onto actual behavior.

Although the four studies described above give consistent evidence that people are more likely to mispredict their own behavior in moral situations than their peers, they also present a puzzle. People have a lifetime of information and experience with themselves, so why do they have more difficulty accurately predicting themselves than others?

One reason that we investigated in the remaining studies focuses on the kind of information people consider when predicting their own and others' behavior. As Kahneman and Tversky (1979) noted in their theoretical analysis of behavioral forecasting (see also Buehler, Griffin, & Ross, 1994), people generally have two types of information on which to base predictions of future behavior: case based and distributional. Case-based information involves evidence relevant to the specific case or person under consideration, whereas distributional information involves evidence about the distribution of behavior in similar or past situations (i.e., base rates of one's own or others' behavior). Of course, in any prediction of oneself and others, people should base their predictions on a combination of the case-based and distributional evidence they have at their disposal. Several analyses demonstrate that people are much more accurate when they do so (e.g., Dunning, Griffin, Milojkovic, & Ross, 1990; Kahneman & Tversky, 1973; Vallone, Griffin, Lin, & Ross, 1990).

One would think that accuracy in prediction might therefore be a simple task, particularly given that people have rather competent intuitions about the distribution of social behavior in a number of domains. People seem to automatically and accurately encode the frequency with which behaviors occur, even when they are not instructed to do so (Hasher & Zacks, 1984; Hasher, Zacks, Rose, & Sanft, 1987; Zacks, Hasher, & Sanft, 1982), and can describe the distributions of social behaviors with considerable accuracy. To be sure, perceptions of social distributions are accompanied by some

biases (e.g., false consensus), but on the whole people accurately report the majority opinion and behaviors of their peers, as well as the shape of the distribution underlying those opinions and behaviors (Nisbett & Kunda, 1985). In a sense, the four studies described above provided even more evidence for this competence in anticipating base rates, at least when predicting others.

Given people's accuracy in perceiving, learning, and reporting distributional information, why do they fail to apply this information to predictions of themselves? Why do people refuse to apply the wisdom gained while observing themselves and others to predictions of their own behavior? The answer appears to lie in the preference people have for basing their predictions on case-based information when they have it (Buehler et al., 1994; Nisbett & Borgida, 1975). When considering ourselves, as C. S. Lewis put it, "We have inside information. We are in the know" (1952, p. 33). As a result, people may base predictions about their own behavior not on the base rate evidence they have accumulated about people in general, but rather on case-based information they possess about themselves, adopting what Kahneman and Tversky (1979) called an "internal approach to prediction." People may assess how they would behave in a moral dilemma by assessing their personality or "character," rather than considering how they have behaved in the past or how they have observed others behaving in similar situations, leading them to make less accurate predictions about themselves than they might otherwise.

Predicting others' behavior, however, is another matter. The notion of "people in general," after all, is a more abstract entity about which very little case-based information may be available. Relatively few concrete images may come to mind when we think of the "average American," the "average woman," or in our participants' case, "the average Cornell student." Consequently, people may more naturally adopt an "external approach to prediction" (Kahneman & Tversky, 1979) when forecasting others' behavior simply because little or no case-based information is available (see also Buehler et al., 1994). Predictions about others' moral behavior might therefore be more strongly based on an assessment of how others have behaved in similar, or occasionally identical, situations in the past. As a result, predictions of others may be more accurate not because of the information people possess about others, but rather because of the information they lack.

This difference in the approach people take in self- versus social prediction suggests that the pattern we observed in the first three studies—more inaccurate predictions of self relative to others—follows from a fundamental and consistent difference in the information people have, and pay attention to, when predicting themselves versus others. Except in rare circumstances, people have case-based information about the self and are thus inclined to ignore distributional information. For predictions about their peers, people have less case-based information and thus give greater weight to distributional information and end up making more accurate predictions.

Study 5A: Practice Makes Perfect?

We investigated this hypothesis more carefully in the remaining two studies. If people base their self-predictions on case-based information and neglect distributional information, then they should be insensitive to distributional information when it is provided to them, even in situations when they competently learn it

and when it is clearly relevant for prediction. When predicting the behavior of their peers, in contrast, they should be willing to give this information more weight and be more inclined to revise their predictions in light of this evidence.

To address this hypothesis in Study 5A, we described the procedures and gradually revealed the results of the donation experiment described earlier (Study 3) and asked people to predict their own and others' behavior. We expected that participants would give more weight to the results of that experiment (i.e., to the distributional information) when predicting their peers than when predicting themselves. Even though the data we gave participants was the most accurate information we had about how people actually respond in this donation situation, we suspected people would apply this knowledge only to their predictions of others.

Method

Participants

A total of 121 Cornell undergraduates participated for extra credit in their psychology or human development courses.

Procedure

On arrival to the lab, participants were informed that this experiment would involve making some predictions about their own and others' behavior. They were then escorted to a private cubicle containing a Macintosh computer and began working through a computer program that contained the same experimental description provided to participants in the distant condition of Study 3. Participants were then asked to make two predictions: how much of their own fee, if any, they would donate if they were actually confronted with the situation, and how much a person selected at random from a sample of 100 Cornell students would donate.

After this first round of predictions, participants were informed by the computer that behavior in novel situations can often be difficult to predict and that their prediction of either their own or others' behavior could be

inaccurate. To aid in these predictions, participants were informed that they would be randomly presented with the actual donations of 13 participants who had experienced this exact situation earlier in the semester and that they could revise either of their initial predictions as they deemed appropriate. Each participant was given three opportunities to revise his or her initial predictions—after being shown the behavior (chosen at random) of 3, 7, and all 13 of the participants in the actual behavior condition of Study 3.

Results and Discussion

We predicted that participants would be willing to revise their predictions of others' behavior in light of the distributional evidence we provided, leading them to make increasingly more accurate predictions, but that they would be reluctant to do the same when predicting their own behavior. To test these hypotheses, we submitted participants' predictions to a 2 (predicting: self or others) × 4 (amount of distributional information: none, 3, 7, or 13 data points) repeated-measures ANOVA. This analysis revealed two interesting effects. First, as seen in Figure 3, participants on average claimed that they would donate more overall ($M = \$2.84$) than would their peers ($M = \$1.93$), $F(1, 120) = 35.49, p < .001$.

Second, this main effect was qualified by the predicted interaction, suggesting that the distributional information had a differential impact on people's predictions of their own and their peer's behavior, $F(3, 360) = 30.86, p < .001$. When predicting their peers, participants became increasingly more cynical with each round of new information, $F(3, 360) = 34.85, p < .001$. Participants initially overestimated the amount that others would actually donate ($M = \$2.20$) but repeatedly revised that prediction to a final estimate that was much closer ($M = \$1.66$) to the actual amount donated in the original study ($M = \$1.53$). When predicting themselves, however, participants felt no need to revise their initial estimates, and receiving increasing amounts of distributional evidence had no impact on their self-predictions, $F(3, 360) = 1.71$,
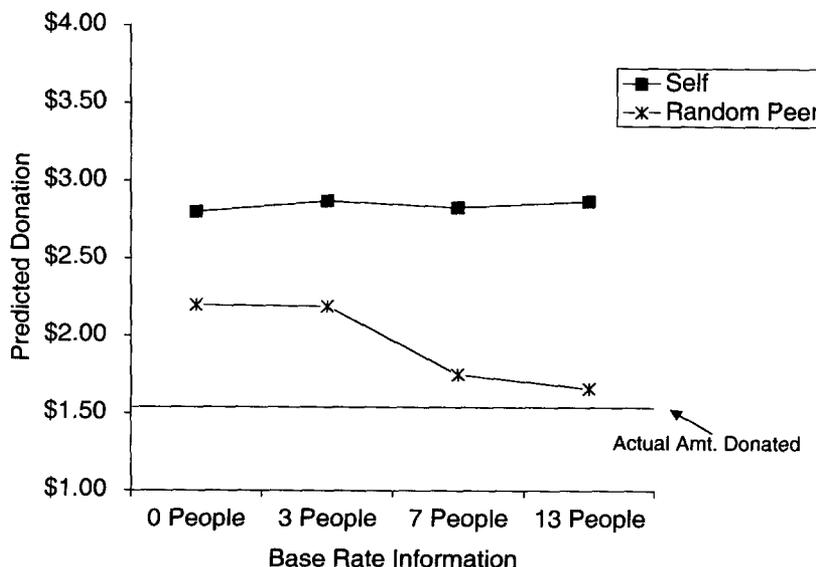

Figure 3. Predicted amount (amt.) of donation for oneself and a random peer after learning the behavior of 0, 3, 7, or all 13 participants in a previous sample (Study 5A).

$p > .15$. If anything, people thought they would be somewhat more charitable after learning how stingy their peers had been.

These results provide initial evidence that the divergence in accuracy of self- and social predictions may result from the different kinds of information people consider when making those predictions. When considering oneself, people may turn inward and focus on case-based information to generate a prediction and be unwilling to consider distributional evidence that would otherwise calibrate this behavioral forecast. Ironically, providing information about how others actually behaved in this experiment did not induce our participants to doubt their lofty self-predictions, but rather led them to feel even more "holier than thou."

Study 5B: When Less (Information) Is More

We argue that the resistance of self-prediction to diagnostic distributional information is a product of the preference for case-based information. However, there are alternative explanations for the resistance we observed in Study 5. Perhaps participants wished to maintain positive images of themselves and so were motivated to disparage the relevance of the distributional information we presented them (Buehler, Griffin, & MacDonald, 1997; Ditto & Lopez, 1992; Kunda, 1987, 1990). Alternatively, perhaps people felt more committed to an initial prediction about the self, and revising a self-prediction might require admitting a potentially embarrassing shortcoming in self-knowledge.

We conducted Study 5B to more conclusively demonstrate that it is the presence of case-based information that leads people to ignore valid distributional information in self-prediction. According to our analysis, any case-based information should prompt people to ignore distributional evidence. To address this possibility we replicated Study 5A but added a third prediction for people to consider. In addition to predicting their own and a random peer's behavior, we gave participants some information about one of their peers and asked them to predict how that person would respond to the opportunity to donate. We expected this individuating information would prompt participants to ignore the distributional information we provided just as participants had done in their self-predictions.

Method

Participants

A total of 74 Cornell undergraduates participated for extra credit in their psychology or human development courses.

Procedure

The procedure in this experiment was closely modeled after that of Study 5A, with a few notable exceptions. First, after being placed in a private cubicle with a computer, participants were given a background survey that asked them to list five words that most accurately described their personality and to write a short paragraph detailing "who you are." Once finished, the experimenter collected the surveys and randomly distributed one to each of the participants in the present or a subsequent session. Participants were then told that they would have an experiment described to them and be asked to make three predictions: how they themselves would respond in the experiment, how the specific person described in the background survey would respond, and how a random Cornell student (selected from a sample of 100) would respond. The order

of these questions was counterbalanced with a Latin-square design but did not influence any of the results.

The donation experiment was then described, and participants predicted how much money they, the specific student they received information about, and a randomly selected Cornell student would donate. They were then presented information about how a random selection of 3, 7, and all 13 participants had responded in the original experiment. After each round of information, participants were given the opportunity to revise any of their previous predictions.

At the end of the experiment participants rated the extent to which they considered case-based and distributional information when making their predictions. For predictions about their own behavior, participants indicated the extent to which they considered "an assessment of their personality" and the "behavior of Cornell students in the original experiment" when making their predictions. Participants made equivalent ratings for their predictions of the individuated and random peer.

Results and Discussion

As before, we expected participants would feel "holier than thou," in this case anticipating that they would donate more money to charity. Our main prediction in this experiment, however, centered on participants' reactions to receiving distributional information. Because the participants possessed case-based information about themselves as well as the individuated peer, we expected they would use this information when making their predictions and be unwilling to revise them in light of the distributional evidence we provided. In contrast, we expected people to be more sensitive to distributional evidence when predicting the behavior of a random peer, gradually becoming more and more accurate as information became available.

To assess these hypotheses, predictions were submitted to a 3 (prediction target: self, individuated peer, random peer) $\times$ 4 (amount of distributional information: none, 3, 7, 13 data points) within-participant ANOVA. Both main effects in this analysis were significant, indicating that participants revised their predictions after learning actual base rates, $F(3, 219) = 16.63, p < .01$, and that predicted donations varied across the three targets, $F(2, 146) = 25.40, p < .01$.

As predicted, these main effects were qualified by a significant interaction, $F(6, 438) = 3.84, p < .01$, indicating that the extent to which people revised their predictions depended on who they were predicting. As can be seen in Figure 4, participants were more willing to revise their predictions of a random peer in light of actual base rates than predictions of either themselves or of the individuated peer, $Fs(3, 219) = 5.51$ and 3.77, respectively, $ps < .05$. As expected, there was no significant difference in the extent to which participants revised predictions of their own and the individuated peer's behavior, $F(3, 219) = 1.28, p > .25$. In the absence of case-based or individuating information, participants were fully willing to revise their initial predictions, becoming increasingly more accurate as base rates became available.

This analysis is bolstered by participants' self-reported consideration of case-based and distributional evidence. As can be seen in Figure 5, planned contrasts in a repeated measures ANOVA revealed that participants reported using "the behavior of Cornell students in the original experiment" more heavily when predicting the random peer's behavior than when considering either their own, $F(1, 73) = 70.53, p < .0001$, or the individuated peer's behavior, $F(1, 73) = 35.83, p < .0001$. Participants also reported
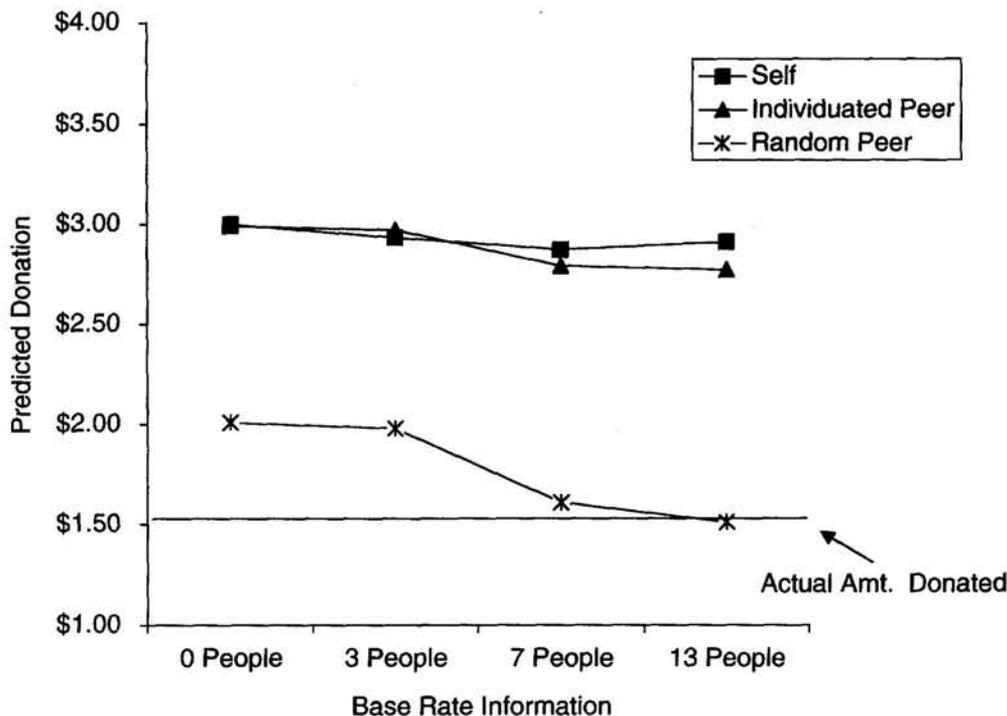
Figure 4. Predicted amount (amt.) of donation for oneself, the individuated peer, and the random peer after learning the behavior of 0, 3, 7, or all 13 participants in a previous sample (Study 5B).

using this factor more heavily when predicting the individuated peer than when predicting themselves, $F(1, 73) = 4.13, p < .05$, but these two figures were much closer than their ratings for the random peer. In contrast, participants reported that they were significantly less likely to consider an assessment of personality when predicting the random peer than when predicting either their own, $F(1, 73) = 83.26, p < .0001$, or the individuated peer's behavior, $F(1, 73) = 80.50, p < .0001$. Participants also reported using this information less when predicting the individuated peer's behavior than when predicting their own, $F(1, 73) = 7.34, p < .01$, but again these two figures were more similar than reports for the random peer.

One somewhat unexpected finding in this experiment was the degree to which predictions of an individuated peer mapped onto self-predictions, not only in terms of the predicted insensitivity to

Figure 5. Self-reported consideration of personality assessment and base rates on predictions for oneself, the individuated peer, and the random peer (Study 5B).

base rates but also in terms of the anticipated charitability of behavior. Indeed, participants felt "holier than thou" only in relation to the random peer, $F(1, 73) = 40.68, p < .001$, not in relation to the individualized peer $(F < 1, ns)$.

We suspect this latter effect occurred for two reasons. First, research suggests that specified others are rated as above average on a variety of dimensions relevant to moral dilemmas, including friendliness, fairness, consideration, politeness, and generosity (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Klar & Giladi, 1997; Sears, 1983). It is likely that these trait ratings would map directly onto behavioral predictions, leading to more favorable predictions of individuated than random peers. Second, the individuating information participants received was a self-description provided by another participant. Given people's well-documented tendency to think of themselves in the best possible light (Alicke, 1985; Dunning et al., 1989; Weinstein, 1980), this individuating information was more likely to imply generosity than selfishness. We doubt participants would anticipate such admirable behavior from someone who described themselves as egotistical, selfish, and greedy, but none of our participants chose these traits to describe themselves.

General Discussion

> When Thales was asked what was difficult, he said, "To know one's self." And what was easy, "To advise another."
> —Diogenes Laertius, *Lives of Eminent Philosophers*

The findings reported in this article provide consistent support for Thales's lament. Across four studies, participants showed considerable difficulty accurately predicting their own behavior, consistently anticipating they would behave more morally than data suggested they would. Participants overestimated how likely
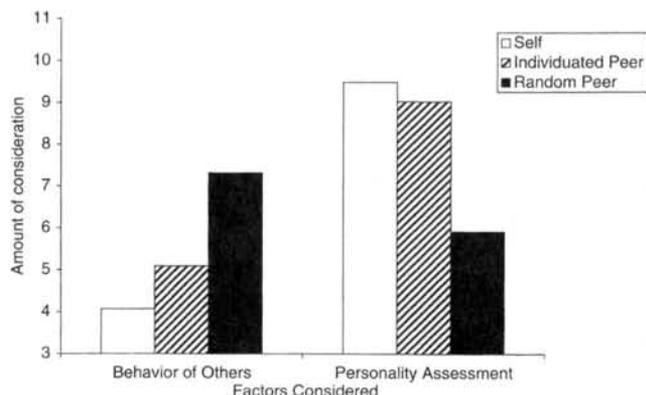
they were to buy a daffodil to support the American Cancer Society (Study 1), to cooperate in a prisoner's dilemma (Study 2), to donate part of an experimental participation fee to charity (Study 3), and to shackle themselves rather than a partner to a longer and more unpleasant experimental task (Study 4). They also overestimated how sensitive their behavior would be to moral sentiments and underestimated how responsive they would be to self-interest concerns (Study 4). In all cases, participants' predictions of their peers were more accurate. Taken together, these findings suggest that participants, in moral dilemmas at least, may well have more difficulty "knowing themselves" than "advising others."

In addition to validating Thales' observation, Studies 5A and 5B suggest that this difference in accuracy may be a basic feature of self- and social predictions, produced by the different perspectives people adopt when making these predictions. When considering the behavior of other people in general, participants showed a willingness to base their predictions on distributional information that resulted in more accurate predictions. However, when predicting their own behavior, participants were reluctant to consider this information and continued to make unrealistically moral predictions. Although they were given increasing amounts of information about how other people react to the situation under consideration, participants continued to persist in the belief that they would act in kind and generous ways. This difference makes the divergence in accuracy of self- and social predictions more understandable: People virtually always have case-based information about themselves but not others, leading them to base their predictions of others on more diagnostic distributional evidence.

Study 5B provided further evidence that it is the presence of case-based evidence, and not a motivated desire to maintain a positive self-image, that produced the resistance of self-prediction to distributional information. When participants were given individuating information about another person, they tended to ignore distributional information in their predictions of that person's behavior. Regardless of how much they were told about how other students reacted to an opportunity to contribute to charity, participants were relatively reluctant to revise their predictions of someone they "knew something about."

There are, however, two caveats suggested by our data on the case-based nature of self-prediction that deserve mention. First, although Study 5B ruled out a motivational explanation for the flattering predictions of one's own moral behavior, it does not demonstrate that motivation is out of the picture altogether. The desire to think well of oneself may have produced the rather favorable case-based materials that the participants used to make their self-predictions. Over the years, our participants may have interpreted information about themselves in any number of flattering ways (see Baumeister & Newman, 1994; Dunning, in press; Kunda, 1990, for reviews), leaving them with a rather favorable impression of themselves. Once in place, these favorable case-based materials could have led them to mispredict how they would respond to moral dilemmas, even if they were trying to be as accurate as possible and to impartially review what they believed about themselves.

Second, our analysis does not demonstrate that people will always be more accurate when predicting others' behavior. In particular, our data, and that of previous researchers (e.g., Nisbett & Kunda, 1985), suggests that people will give relatively accurate predictions about how other people in general will behave, as well as any individual for whom they fail to have case-based information. As soon as people learn some specific and concrete information about another person, then their predictions, like those made about the self, may be prone to inaccuracy. In addition, this means that the predictions people make about individuals they know and think highly of, such as their significant others or family members, may show as much inaccuracy as self-predictions. To the extent that people possess case-based information about another person, they are likely to ignore valid distributional information that they can estimate, learn, and ultimately use to fashion more accurate predictions.

Limitations?

At this point a critical reader may begin to wonder about further limitations to the patterns of self- and social predictions we found in this experiment. In particular, in all of the experiments we report, the actual base rates of charitable behavior were somewhat low, or at least considerably less than 100%. Might the pattern we found reverse itself when actual base rates are much higher, with people predicting their own behavior more accurately and underestimating how ethically their peers might behave?

Although we cannot say with certainty whether such a pattern might arise, we believe several features of our analysis argue against it. First, it was our intention in this research to explore people's intuitions about themselves and others when they were placed in situations that presented a real conflict between moral sentiments and self-interest. When conflict between these two forces arose, which would win out? And would people correctly anticipate this outcome? Situations with this kind of conflict, by definition, produce moderate base rates of moral behavior. If, instead, people always performed the moral action, or refused to perform it, there would be no evidence of real conflict, and we deliberately avoided situations that would produce extremely high or low base rates of moral behavior.

Despite this deliberate avoidance, there was substantial variability in the base rates of moral behavior among the experiments we did run, from 13% (in one condition of Study 4) to 62% (Study 3). In each case, the basic pattern of accuracy was replicated: People grossly overestimated the rate at which they themselves would act charitably but were relatively more accurate in their predictions of others. Furthermore, participants in Study 4 proved more accurate in predicting how situational variations would change the base rates of moral behavior, at least for their peers. People tracked changes accurately in our experiments when base rates were low, and we doubt participants would suddenly lose this ability when base rates were high.

Second, when base rates for moral behaviors become extremely high and the conflict between self-interest and moral sentiments becomes moot, people might not necessarily anticipate a difference between their own and others' behavior. For instance, people probably do not feel much more likely than their peers to stop at a red light, pay a cashier for their groceries, or behave compassionately toward a distraught infant. For these reasons, we doubt that the pattern of accuracy we documented would reverse itself with extremely high rates of moral behavior.

Alternative Interpretations

We interpret the results of these studies as indicating, in the moral domain at least, that people mispredict themselves more than others and that these predictions are produced by the divergent perspectives people adopt when predicting behavior. There is, however, an alternative that requires none of this theoretical discussion, one based solely on social desirability. The discrepancy between participants' predictions for themselves and others, as well as the discrepancies between predicted and actual behavior, may be produced by participants' attempts to appear more moral, and less self-interested, in the eyes of the experimenter than people know themselves to be.

There are three reasons we believe a self-presentation account seems an unlikely explanation for the data we report. First, with the exception of Study 1, great pains were taken to assure participants that their responses were completely anonymous. Participants were informed of this feature numerous times by the experimenter, never asked to write their names on any questionnaire, completed their questionnaires alone in individual cubicles, and returned their responses either in a sealed envelope or anonymously onto a stack of other questionnaires. If presentational concerns were operating, they were operating in service to the self, not the experimenter or anyone else.

Second, although not done explicitly, the Daffodil Days experiment (Study 1) provided something of an implicit manipulation of social desirability but nevertheless produced results similar to the other studies. When predicting as well as when reporting their behavior, participants' responses were not made anonymously (their names were written directly on the questionnaire) and were completed in the observable presence of friends and classmates. Despite this plausible increase in self-presentational concerns, participants provided responses that followed the exact pattern found in subsequent studies. The mean effect size on the difference between predictions of one's own and others' behavior in the Daffodil Days study is nearly identical to the comparable mean effect size in the other four studies (ds = .51 and .48, respectively).

Finally, an explanation based entirely on self-presentational concerns is inconsistent with the results of two of our studies. If people were simply motivated to present the best possible image to the experimenter or themselves, they would have predicted opting for the desirable course of action in Study 4, assigning themselves rather than their partner the long task, regardless of the social circumstances. In contrast, we found that participants were fully willing to admit they would be more selfish when their partner was an average college student than when that partner was a 10-year-old girl. In Study 5B, a general desire to think well of oneself would have left people fully willing to revise their predictions of an individuated peer in a more self-serving fashion, but they did not. Consequently, we believe the results of these experiments are more parsimoniously explained by the theoretical interpretation we have offered than by a simple desire to appear more moral than people believe themselves to be.

Reconciliations

The consistent finding across our studies is that people expect to behave more morally than their peers, and by and large are less accurate when predicting their own behavior than when predicting their peers'. When predicting their own behavior, participants tended to underestimate the influence of material self-interest and overestimate the influence of their moral sentiments. Although more accurate when predicting others, participants also showed a slight tendency to underestimate the influence of self-interest when predicting others' behavior. This finding, however, seems inconsistent with recent conclusions from the social psychological literature that people generally overestimate the impact of self-interest on the behavior and attitudes of others (Diekmann, Samuels, Ross, & Bazerman, 1997; Miller, 1999). For example, people think monetary incentives will have more influence on people's willingness to donate blood than it actually does, and that a person's smoking habits will be a better predictor of his or her willingness to endorse smoking-related policies than they actually are (Miller & Ratner, 1998).

Is there any way to resolve this contradiction between our studies and this previous research? We suspect there is, and that whether people overestimate the impact of self-interest or predict it correctly will depend largely on whether the predicted situation involves hedonic consequences that are immediate or remote. In our studies, hedonic consequences were immediate. Participants needed to decide whether they were going to donate money or time at the very moment we asked, not at some distant time in the future. In these cases, self-interest does significantly influence behavior (Green & Cowden, 1992; Regan & Fazio, 1977) and people's cynical beliefs about others are likely to be more well-informed than mistaken.

When hedonic concerns are remote, however, self-interest is less influential and people are likely to overestimate its impact. Predicting whether you would be willing to donate blood at some time in the future (Miller & Ratner, 1998, Study 1), for example, is considerably different from actually doing so when staring down the business end of a needle. Endorsing some antismoking policy that may or may not be enacted at some point in the future is different from handing over one's pack of cigarettes to an agent of the enforcement police. We suspect that people are more likely to overestimate the impact of self-interest when such concerns are remote but not when they are more immediate. The problem with people's beliefs about the power of self-interest, then, may not be that they are systematically wrong, but rather that they are overapplied.

Conclusions

The main goal of this research was to determine which error led people to feel "holier than thou." We found that, in the moral domain at least, people were much less accurate when predicting their own behavior than when predicting others'. Participants consistently, and grossly, overestimated the likelihood that they themselves would act in a selfless and altruistic manner, whereas the predictions made about others tended to converge more closely with reality.

Although we have restricted the current analysis to domains with moral or altruistic overtones, divergences in predictions of oneself and others occur in other areas as well (e.g., Alicke, 1985; Brown, 1986; Buehler et al., 1994; Dunning et al., 1989; Klein & Kunda, 1993; Kruger & Dunning, 1999; McFarland & Miller,

1990; Rothman, Klein, & Weinstein, 1996). We suspect that the basic mechanisms we have outlined here would hold in these other domains as well and that predictions about emotional reactions to a happy or sad event (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Schkade & Kahneman, 1998), likely completion time for a difficult project (Buehler et al., 1994), and probability of breaking up with a love-smitten partner (Weinstein, 1980) will exhibit the same patterns of accuracy for social predictions and inaccuracy for self-predictions we observed in these experiments. Still, it is possible that moderating variables could attenuate our findings and that they may not generalize to all contexts. Identifying potential moderators and exploring more generally the accuracy of self- and social predictions would likely be a fruitful avenue of future research.

But whatever future research reveals about the accuracy of self- and social prediction, we might be well served to start this venture by letting the thinkers of old guide our hypotheses. The results presented in this article were anticipated not only by Thales but also by scholars of other cultures as well. Self-knowledge, after all, has long been considered a virtue, and not an easy one to attain. As the Chinese philosopher Lao Tzu observed, "He who knows others is learned. He who knows himself is enlightened" (Chan, 1963, p. 159).

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621–1630.

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

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter than others: The Muhammad Ali effect. *Social Cognition, 7,* 275–295.

Baumeister, R. F., & Newman, L. S. (1994). Self-regulation of cognitive inference and decision processes. *Personality and Social Psychology Bulletin, 20,* 3–19.

Berke, R. L. (1998, February 15). Clinton's o.k. in the polls, right? *The New York Times,* pp. 4:1, 4:5.

Bierbrauer, G. (1976). Why did he do it? Attribution of obedience and the phenomenon of dispositional bias. *European Journal of Social Psychology, 9,* 67–84.

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social judgments. *Social Cognition, 4,* 353–376.

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin, 23,* 238–247.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67,* 366–381.

Chan, W. (1963). *The way of Lao Tzu.* Indianapolis, IN: Bobbs-Merrill.

Diekmann, K. A., Samuels, S. M., Ross, L., & Bazerman, M. H. (1997). Self-interest and fairness in problems of resource allocation: Allocators versus recipients. *Journal of Personality and Social Psychology, 72,* 1061–1074.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology, 63,* 568–584.

Dunning, D. (in press). On the motives underlying social cognition. In N. Schwarz & A. Tesser (Eds.), *Blackwell handbook of social psychology: Vol. 1. Intrapersonal processes.* New York: Blackwell.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58,* 568–581.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75,* 617–638.

Goethals, G. R. (1986). Fabricating and ignoring social reality: Self-serving estimates of consensus. In J. M. Olson, C. P. Herman, & M. P. Zanna (Eds.), *Social comparison and relative deprivation: The Ontario symposium* (Vol. 4, pp. 135–157). Hillsdale, NJ: Erlbaum.

Goethals, G. R., Messick, D. M., & Allison, S. T. (1991). The uniqueness bias: Studies of constructive social comparison. In J. Suls & T. A. Will (Eds.), *Social comparison: Contemporary theory and research* (pp. 149–176). Hillsdale, NJ: Erlbaum.

Green, D. P., & Cowden, J. A. (1992). Who protests: Self-interest and white opposition to busing. *Journal of Politics, 54,* 471–496.

Greenwald, A. G., Carnot, C. G., Beach, R., & Young, B. (1987). Increasing voting behavior by asking people if they expect to vote. *Journal of Applied Psychology, 72,* 315–318.

Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality and Social Psychology, 59,* 1128–1139.

Griffin, D. W., & Ross, L. (1991). Subjective construal, social inference, and human misunderstanding. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 319–359). San Diego, CA: Academic Press.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39,* 1372–1388.

Hasher, L., Zacks, R. T., Rose, K. C., & Sanft, H. (1987). Truly incidental encoding of frequency information. *American Journal of Psychology, 100,* 69–91.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 313–327.

Kahneman, D., & Tversky, A. (1979). Intuitive prediction: Biases and corrective procedures. *TIMS Studies in Management Science, 12,* 313–327.

Klar, Y., & Giladi, E. E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology, 73,* 885–901.

Klein, W. M., & Kunda, Z. (1993). Maintaining self-serving social comparisons: Biased reconstruction of one's past behaviors. *Personality and Social Psychology Bulletin, 19,* 732–739.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121–1134.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53,* 37–54.

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

Lewis, C. S. (1952). *Mere Christianity.* New York: Macmillan.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes, 65,* 272–292.

Loewenstein, G., & Adler, D. (1995). A bias in the prediction of tastes. *Economic Journal, 105,* 929–937.

McFarland, C., & Miller, D. T. (1990). Judgments of self–other similarity:

Just like other people, only more so. *Personality and Social Psychology Bulletin, 16,* 475–484.

Miller, D. T. (1999). The norm of self-interest. *American Psychologist, 54,* 1053–1060.

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology, 74,* 53–62.

Nisbett, R. E., & Borgida, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology, 32,* 932–943.

Nisbett, R. E., & Kunda, Z. (1985). Perception of social distributions. *Journal of Personality and Social Psychology, 48,* 297–311.

Regan, D. R., & Fazio, R. (1977). On the consistency between attitudes and behavior: Look to the method of attitude formation. *Journal of Experimental Social Psychology, 13,* 28–45.

Ross, L., & Ward, A. (1995). Psychological barriers to dispute resolution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 27, pp. 255–304). San Diego, CA: Academic Press.

Rothman, A. J., Klein, W. M., & Weinstein, N. D. (1996). Absolute and relative biases in estimations of personal risk. *Journal of Applied Social Psychology, 26,* 1213–1236.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9,* 340–346.

Sears, D. O. (1983). The person-positivity bias. *Journal of Personality and Social Psychology, 44,* 233–250.

Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology, 39,* 211–221.

Smith, A. (1976). *A theory of moral sentiments.* Oxford, England: Clarendon Press. (Original work published in 1759)

Vallone, R. P., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident prediction of future actions and outcomes by self and others. *Journal of Personality and Social Psychology, 58,* 582–592.

Van Boven, L., Dunning, D., & Loewenstein, G. (2000). Egocentric empathy gaps between owners and buyers: Misperceptions of the endowment effect. *Journal of Personality and Social Psychology, 79,* 66–76.

van Lange, P. A. (1991). Being better but not smarter than others: The Muhammad Ali effect at work in interpersonal situations. *Personality and Social Psychology Bulletin, 17,* 689–693.

van Lange, P. A., & Sedikides, C. (1998). Being more honest but not necessarily more intelligent than others: Generality and explanations for the Muhammad Ali effect. *European Journal of Social Psychology, 28,* 675–680.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820.

Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic encoding of event frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 106–110.