

Correct acceptance weighs more than correct rejection: a decision bias induced by question framing

Yaakov Kareev · Yaacov Trope

Published online: 6 November 2010
© Psychonomic Society, Inc. 2010

Abstract We propose that in attempting to detect whether an effect exists or not, people set their decision criterion so as to increase the number of hits and decrease the number of misses, at the cost of increasing false alarms and decreasing correct rejections. As a result, we argue, if one of two complementary events is framed as the positive response to a question and the other as the negative response, people will tend to predict the former more often than the latter. Performance in a prediction task with symmetric payoffs and equal base rates supported our proposal. Positive responses were indeed more prevalent than negative responses, irrespective of the phrasing of the question. The bias, slight but consistent and significant, was evident from early in a session and then remained unchanged to the end. A regression analysis revealed that, in addition, individuals' decision criteria reflected their learning experiences, with the weight of hits being greater than that of correct rejections.

Keywords Decision making · Type I and Type II errors

"Will it rain tomorrow?" "Is the defendant guilty?" "Is the tumor malignant?" Every one of these questions implies an

effect of interest—the raining, the guilt of a defendant, the malignancy of a tumor. We propose that because people are intent on detecting effects (Kareev & Fiedler, 2006; Kruglanski, 1989, 1990; Lopes, 1982; Trope & Liberman, 1996), they will set the decision criterion for answering such questions in a way that would increase the chances of detecting an effect when one exists, and decrease the chances of realizing that there is none. With positive-test strategy being common place (e.g., Klayman & Ha, 1987; Trope & Bassok, 1982), we assume that the phrasing of a question indicates the effect of interest, and implies what combination of answer and outcome constitutes a hit, a false alarm, a miss, or a correct rejection. Therefore, if our claim is correct, the complementary phrasing—"Will it not rain tomorrow?" "Is the defendant innocent?" "Is the tumor benign?"—would result in a different setting of the decision criterion.

Decision theory (e.g., Green & Swets, 1966) identifies three factors that ought to be considered when determining a decision criterion: The separability of the distributions that generate positive and negative items, the base rate of each type of item, and, for either answer, the payoffs associated with being correct or incorrect. With decision theory providing a normative benchmark, empirical research involving signal detection or probabilistic categorization tasks revealed that all three factors indeed affect behavior in a manner commensurate with the theoretical analysis. At the same time, the observed value of the decision criterion has at times deviated from the value prescribed by decision theory. Factors that were found to cause such deviations included improper weighting of base-rates and payoffs (Healy & Kubovy, 1981; Kubovy & Healy, 1977), valuing response accuracy beyond the value specified in the payoffs (Bohil & Maddox, 2001; Maddox, 2002; Maddox & Bohil, 1998), and one's unique learning experience while performing the task (Erev, 1998).

Electronic supplementary material The online version of this article (doi:10.3758/s13423-010-0019-z) contains supplementary material, which is available to authorized users.

Y. Kareev (✉)
School of Education and The Center for the Study of Rationality,
The Hebrew University of Jerusalem,
Jerusalem 91904, Israel
e-mail: kareev@vms.huji.ac.il

Y. Trope
New York University,
New York, NY, USA

We propose that, beyond the effects of all such factors, the way a question is phrased produces a systematic bias for detecting the alternative designated as focal by the phrasing. Since such a bias would amplify the payoffs associated with hits and misses, and attenuate those of false alarms and correct rejections, people would lower the decision criterion for assigning an item into the focal category. As a result, they would answer in the affirmative more often than theoretically expected given the payoffs specified by the experimenter. In other words, we hypothesize an asymmetry that favors predicting the focal alternative when the payoff structure is symmetric. If we are correct, and the quest for effects results in subjective payoffs that are different from those externally defined, the proportion of positive answers would be higher than that expected on normative grounds.

Our claim that the phrasing of a question could affect the setting of a decision criterion sits well with the work on framing effects by McKenzie and his co-workers (McKenzie, 2004; McKenzie & Nelson, 2003; Sher & McKenzie, 2006). McKenzie and his co-workers point out that logically equivalent phrasings may not be information-equivalent; that normatively relevant information can be leaked via the experimenter's choice among logically equivalent frames. Consistent with this conceptualization, we propose that information about an effect that is of interest to a person posing a question leaks, is revealed by, the phrasing of the question. The phrasing thus provides subtle cues about the payoff structure that may go beyond those explicitly stated. However, it should be noted that McKenzie's own work (2004; Sher & McKenzie, 2006), as well as his reinterpretation of other work involving the detection of covariation and hypothesis testing (e.g., McKenzie & Mikkelsen, 2000, 2007; Oaksford & Chater, 1994, 1996), suggests that the phrasing chosen in everyday tasks indicates the rare, rather than the more common values. In contrast, we hypothesize that the phrasing is taken to indicate the value of interest, leading people to act as if that value is more common, not rarer. Thus, the convention observed by McKenzie works against our hypothesis.

There is prior evidence that people do tend to respond to Yes/No questions in the affirmative. The acquiescence bias—a tendency to respond in the affirmative to a question asked (e.g., Cronbach, 1942; Hodgins & Zuckerman, 1993; Jackson & Messick, 1958; Knowles & Condon, 1999; Messick & Jackson, 1961; Zuckerman, Knee, Hodgins, & Miyake, 1995)—is one type of such evidence. However, this bias has been observed in situations in which base rates and payoffs were neither assessed nor controlled. In fact, a phrasing-induced bias, if we show it to exist, could explain the acquiescence bias. Referring to a similar tendency, Gilbert (1991) pointed out that everyday experiences predispose people to believe that propositions are true and

hence to respond to questions in the affirmative. Gilbert's explanation is orthogonal to ours, but the data he cites could well reflect the subtle bias we suggest.

Most relevant for our argument is the work of Wallsten and his co-workers (Wallsten, Bender, & Li, 1999; Wallsten & González-Vallejo, 1994). Their participants observed statements like "In 1990, the population of Michigan exceeded that of Minnesota", and were asked to judge whether they were true or false. Base rates were controlled by using complementary statements, and explicit payoffs were associated with each outcome. Most participants (more so in Wallsten & González-Vallejo, 1994, but also in Wallsten et al., 1999) exhibited a bias to respond "true" under symmetric payoffs. In fact, Wallsten et al. (1999) conclude that their findings are compatible with the idea that the subjective cost of calling a true statement false (i.e., a miss) exceeds the subjective cost of calling a false statement true (i.e., a false alarm).

Suggestive as they are, these earlier observations do not provide an unequivocal answer to our question. The acquiescence bias could reflect the payoffs implied by a question-asker's choice of phrasing, but in the situations it has been observed neither the payoff structure, nor the base rates were specified. Gilbert's arguments relate only to base rates, with no consideration of payoffs. Finally, because participants in the studies by Wallsten and his co-workers received no feedback, these studies provide no evidence regarding the effects of different outcomes on the eventual setting of the decision criterion.

To test our claim, we set up the following prediction task: Every participant was presented with a full deck of regular playing cards, face up, and asked to answer, for each card, the same yes/no question about the marking on its back (e.g., "Does it have an X on its back?"). In some decks marking involved the presence or the absence of an X; in others it involved a pair of letters that could either be the same or different. Correct answers were rewarded and incorrect answers were penalized. Payoffs were symmetrical—the rewards for the two types of correct answers were the same, and so were the penalties for the two incorrect answers. Base rates (at .5/.5) and the relationship between the face of a card and the marking on its back (none) were not announced in advance. For generalization, we used four different questions which involved the following values: "X?", "Not X?", "Same marks?", "Different marks?". We asked about both outcomes to control for possible contents effects and included the Same/Different pair of questions to generalize beyond the present/absent (X/Not-X) pair. The overall proportion of "Yes" answers, revealing the setting of the decision criterion, was the main dependent variable.

In our setup—with symmetric payoffs, equal base rates, and zero predictability—decision theory does not prescribe where to set the decision criterion: Since all cutoff points

have the same expected payoff, there would be no reason to favor one over the other. Still, there are two reasons to expect the average criterion to be .50. One reason has to do with individual learning experiences. As base rates had not been announced and the sequence actually observed was random, we expected the decision criterion not only to reflect the decision bias, if one existed, but also to be related to each individual's unique learning experience (see Bohil & Maddox, 2001; Erev, 1998; Healy & Kubovy, 1981; Kubovy & Healy, 1977; Maddox, 2002; Maddox & Bohil, 1998). Because participants were expected to encounter an equal number of Yes and No cards, and to experience an equal proportion of successes and failures, it could be expected that on average the decision criterion would be at .50 if no other effect was at play.

The second reason to expect a decision criterion at .50 is that people tend to balance the regrets experienced by having chosen different alternatives (Selten, Abbink, & Cox, 2005; Selten & Stoecker, 1986; for further empirical evidence see Avrahami, Güth, & Kareev, 2005). If the regret for either incorrect answer—a miss and a false alarm—reflected only the objective difference between actual penalties and would-be rewards, regrets would be balanced with the decision criterion set, on average, at .50. Therefore, any deviation from that value could be taken to indicate a differential weighting of the costs of the two types of errors.

In line with our hypothesis, we expected to find an overall tendency for the criterion to be at a point that would result in more hits, and fewer misses, at the unavoidable cost of fewer correct rejection and more false alarms, with the designation of what constitutes each of these outcomes depending on the phrasing of the question.

Method

We marked the backs of the playing cards of a standard deck. The 52 cards were presented one at a time, face up, and the participants were asked to predict what was on the back of each card. Correct predictions were rewarded, and incorrect predictions penalized. The phrasing of the question, asked before each trial, differed between participants. The participants did not know in advance the prevalence of the focal value, or whether there was any relationship between the face value of the cards and the values on their backs.

Materials

Four decks of cards were employed. In two of them, the back of each card was either marked or not marked with the letter *X*. In the other two decks, each card had two letters on its back—*XX*, *XO*, *OX*, or *OO*. To control for any prior

associations between face value and answer, we marked a given card differently within each pair of decks, so that the correct prediction for it differed. In addition, when marking the cards, we took care to avoid any correlation between color, suit, or value and the mark (or marks) on the back of the cards.¹ A single deck was used for each participant.

The question asked before each card was presented was one of the following four: "Is there an *X* on the back of this card?", "Is there no *X* on the back of this card?", "Are the letters on the back of this card the same?", or "Are the letters on the back of this card different?". For a given participant, the question asked was always the same. In each deck, the correct answer was "yes" for half of the cards and "no" for the other half.

Procedure

Participants were tested individually in a quiet room. While shuffling the cards, the experimenter explained that it was a standard deck consisting of 52 cards and described the mark (or marks) that could appear on the back of each card. She then informed the participant that she would ask the same question about the back of each card, and that the question could be answered by "yes" or "no." Once the question was answered, the card was turned over to check the prediction. Correct answers were rewarded by a gain of 1 new Israeli shekel (NIS; about \$0.25), and incorrect answers were penalized by a loss of 0.5 NIS.

Following the prediction task, participants answered some additional questions related to their performance. Because the focus of this paper is on predictions, we describe these questions and the answers to them in the [supplementary materials](#). The expected earnings in value of participating in the experiment were about 29 NIS (13 NIS in the prediction task and about 16 NIS in the additional tasks).

Participants

Participants were 128 students at the Mt. Scopus campus of the Hebrew University of Jerusalem. They volunteered to participate in the experiment in return for the payment offered. There were 16 students—eight males and eight females—in each of the eight groups defined by the combination of question asked and deck of card used.

¹ To anticipate, the average number of correct responses was 24.98—slightly below the 26 correct responses expected by chance. Clearly, no correlation was identified between the cards and their markings.

Results and discussion

Our main interest was to find out if affirmative responses were more prevalent overall than negative ones. We expected to observe such an overall tendency beyond the effects, if any, of the participants' unique experiences. At the same time, we were also interested to find out if, and in what way, participants' unique experiences, such as the base-rate encountered or the reinforcement history, were related to the proportion of affirmative responses. To that end, we conducted regression analyses that enabled us to assess the relationship between experience and the tendency to respond in the affirmative and, in particular, the weight of different aspects of that experience.

Before reporting the results of these analyses, it should be pointed out that the specific question asked did not have a significant main effect and was never part of a significant interaction. The data were therefore pooled across questions.

Overall bias The average number of "yes" responses was 27.52 (out of 52; i.e., .53 of the cases). This value was significantly larger than 26.00, the number expected if responses corresponded to the average base rate or if responses were random, $t(127) = 2.01$, $p = 0.046$. Thus, as we hypothesized, the phrasing of the question biased responses such that the participants had more hits and fewer misses, at the cost of fewer correct rejections and more false alarms.

If the behavior we observed reflects a deep-seated bias, as our hypothesis implies, it should have been evident early on in the prediction task. To test if this was the case, we conducted several analyses. As an initial test, free of any learning effects, we checked the proportion of "yes" responses in the very first trial.² Seventy of the 128 participants (.55) responded in the affirmative. The prevalence of positive responses on the first trial—above the chance level and close to the overall proportion (.53)—indicates that the overall tendency to respond in the affirmative was evident even when no learning could have taken place.

To test if the prevalence of positive responses changed over time we recorded their proportion in the four quartiles of the session. The proportion of positive responses was .54, .53, .54, and .51 in the first through fourth quartiles, respectively. A test for change over time revealed the linear trend to be nonsignificant, $F(1, 127) = 1.86$, $p = 0.175$, $\eta_p^2 = .014$. We therefore argue that the overall tendency to answer in the affirmative is deeply ingrained, and withstands learning experiences.

The bias observed is in line with that observed by Wallsten et al. (1999), and so is our interpretation, that the subjective cost of a miss exceeds that of a false alarm (or that the subjective value of a hit exceeds that of a correct rejection). If we assume that the decision criterion is set such that regrets for the two types of errors are balanced (see the Impulse-Balance equilibrium notions; e.g., Selten et al., 2005; Selten & Stoecker, 1986), our data suggest that, on average, the subjective cost of a miss is 12% higher than that of a false alarm.

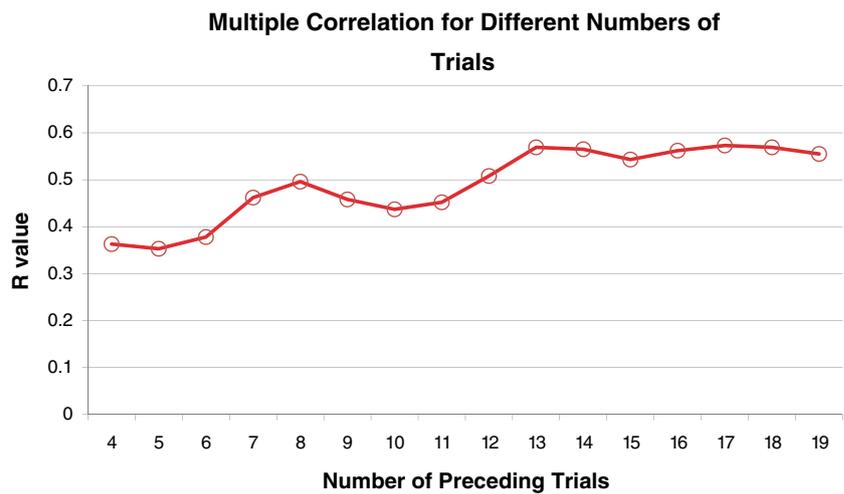
The relationship of positive answers to individual experiences Although we found an overall tendency to respond in the affirmative, we also observed individual differences in the propensity to do so. Wallsten et al. (1999) too report that a majority of their participants, but not all of them, exhibited the bias to respond "true". Could these differences be accounted for by differences in the unique sequence of experiences that the participants had faced? To answer this question, we regressed the overall proportion of positive answers of each participant on three aspects of that participant's initial experience: Base rate (the proportion of positive cases among the cards observed), the proportion of hits, and the proportion of correct rejections.³ In deciding which aspects of the participants' experience to use as predictor variables we had the following considerations: The base rate was deemed important because decision theory prescribes that the setting of the decision criterion should correspond to an estimate of it; the other two variables—hits and correct rejections—represent positive reinforcement for either type of answer (see Erev, 1998, for a successful application of reinforcement learning to probabilistic categorization tasks).

Our analyses focused on the participants' initial experiences because it has been shown that in tasks of repeated decisions which require the estimation of a population parameter, it is initial experiences that are most effective in shaping overall behavior (Hogarth & Einhorn, 1992; Kareev, Aron, & Horwitz-Zeliger, 2002). To find out what segment of initial experiences best predicted overall behavior, we ran 16 separate regression analyses, one for every sequence length between four and 19 trials. The predictors in each analysis were the proportion of positive items, the proportion of hits, and the proportion of correct rejections in the sequence of trials of the specified length. In all the analyses, the criterion to be predicted was the overall proportion of positive responses. The shortest sequence considered consisted of four trials because with fewer trials too many participants

² We thank an anonymous reviewer for suggesting this analysis.

³ Because the proportions of false alarms and misses complement those of hits and correct rejections, the former could have been used in this analysis instead.

Fig. 1 Strength of the multiple correlation: prediction of "yes" responses by the proportions of positive events, hits, and correct rejections as a function of the number of initial trials over which information was accumulated



could not be included in the analysis due to having given only one response; we stopped at a sequence consisting of 19 trials because the analysis indicated that an asymptote had long been reached by then.

The regression analyses revealed that participants' overall proportion of "yes" responses was indeed related their unique experiences (base rates of positive cases encountered and reinforcement history) during the early stages of the session. In the analyses involving only a small number of cards (the first four, five, and six trials), the only significant predictor of the overall proportion of positive responses was the base rate observed. This predictor reached a peak *r* of .38 after six trials. Thus, early on, participants apparently took note only of the base rate of the phenomenon in question and it must have affected their setting of the overall decision criterion. In analyses involving experience over a larger number of trials, the proportions of hits and correct rejections became the strongest predictors of the overall proportion of positive responses. Of the two, the (positive) weight of hits was consistently stronger than the (negative) weight of correct rejections.⁴ The multiple correlation reached a peak of *R* = .57, a value first observed after 13 trials. Once the proportions of successes due to hits and correct rejections was entered into the regression equation, the base rate of positive events had the lowest weight of the three variables, with its residual contribution acting as a moderator. Figure 1 presents the value of the multiple correlation observed after each of the four through 19 first trials.

In light of these findings, it was necessary to check whether the overall tendency to give "yes" responses could have resulted from chance fluctuation in the proportion of positive cases, hits, or correct rejections in the early stages

of the session. This was not the case: The proportion of positive events during the first four, the first five, and the first six trials was slightly below the .50 expected by chance (.46, .47, and .47, respectively). At trials 1 through 13, the proportion of positive events encountered was .49, the proportion of hits (out of "yes" answers) was .47, and that of correct rejections (out of "no" answers) was .48. Thus, it is clear that the overall bias to answer in the affirmative could hardly be accounted for by the early experiences of our participants.

We find three implications of the regression analyses particularly noteworthy: First, the effects of the base rate and reinforcements demonstrate that the participants' responses were strongly related to their initial history of encounters with the environment and their success in handling it. Second, the fact that the regression weight associated with the proportion of hits was consistently stronger than that associated with the proportion of correct

Table 1 The proportion of "yes" responses on trials six and 11, as a function of the proportion of positive items observed in all previous trials

Number of positive items observed	Proportion of positive items observed	Proportion of "yes" responses	Number of participants (<i>n</i>)
In trials 1–5	In trials 1–5	On trial 6	
0 or 1	0.15	0.46	26
2 or 3	0.49	0.58	86
4 or 5	0.86	0.63	16
Overall	0.47	0.56	128
In trials 1–10	In trials 1–10	On trial 11	
0 thru 3	0.24	0.32	22
4 thru 6	0.50	0.57	97
7 thru 10	0.74	0.89	9
Overall	0.47	0.55	128

⁴ For example, following 13 trials, the weights of the proportions of hits and correct rejections were .755 and –.655, respectively.

rejections points to the greater significance of the former in determining behavior. Finally, it may seem impossible to determine whether it was the attraction induced by hits or the repulsion induced by false alarms that affected behavior (the proportion of one being the complement of the other and therefore their weights being opposite in sign but identical in magnitude). Still, we suggest it is the former, rather than the latter, that determined behavior: If responses were driven by false-alarm-induced repulsion, one would expect a preponderance of "no" responses, contrary to the preponderance of "yes" responses observed. Furthermore, a trial-by-trial inspection of behavior revealed no reduction in the propensity to respond in the affirmative following a false alarm, but an increase in the propensity to respond in the affirmative following a miss.

A final point has to be made about our use of the term "bias". We used the term in a neutral sense, and explained why a deviation from an equal number of positive and negative answers would constitute such bias. As used by most, the term connotes sub-normative behavior, and it could be argued that no proportion of "yes" responses would constitute a bias in this, more common, sense because in our study any proportion of "yes" responses had the same expected value, and hence could not be regarded as sub-normative. To justify the use of the term, the argument goes, we should show a propensity to respond in the affirmative across different base rates of positive instances.⁵ Fortunately, the setup of our study was such that it allowed a test of the appropriateness of using the term "bias" in the stronger sense, implying an inappropriate decision procedure: Because the order of cards was random, participants differed with respect to the proportion of positive instances they encountered; we could thus compare the base rate of positive instances they had observed at any point with their prediction for the next trial. We present in Table 1, for illustration, the responses at two points, after five and after ten trials, for the total population as well as for sub-groups that had observed different proportions of positive items prior to the trial in question. As can be seen, the overall proportion of "yes" responses is greater than that previously encountered by the participants. A breakdown of the responses by the proportion of positive cases encountered, reveals the proportion of positive predictions to be above that encountered whenever the proportion encountered was smaller than or (about) equal to .5, and, after ten trials, even when that proportion was above .5. The only exception—participants who observed a high proportion of positive items among the first five—could well represent a "gambler's fallacy", and even there, note that the overall bias up (for the low-proportion group) is stronger than the bias down. Thus, whether bias is defined by failure to

match the base rate or by failure to maximize in view of the base rate, participants could be said to be biased even in the stronger sense of the word.

Conclusions

Taken together, our results demonstrate that in making predictions, people exhibit a slight but consistent bias whose effect is to increase the chances of scoring a hit and avoiding a miss. This bias could not be ascribed to differences in rewards, penalties, or the base rates of outcomes. At the same time, the degree of the bias was shown to be sensitive to environmental characteristics and individual learning history in the early trials. Still, the bias itself, transcending unique personal experience, reveals what we regard as a deep-rooted preparedness for the detection of effects.

Author Note Work reported in this article was supported by Israel Science Foundation Grants 800/04 and 539/07 to Yaakov Kareev and by National Institute of Mental Health Grant R01 MH59030-06A1 to Yaacov Trope. We would like to thank Greg Murphy for very helpful comments on an earlier version.

References

- Avrahami, J., Güth, W., & Kareev, Y. (2005). Games of competition in a stochastic environment. *Theory and Decision*, *59*, 255–294.
- Bohil, C. J., & Maddox, W. T. (2001). Category discriminability, base-rate, and payoff effects in perceptual organization. *Perception & Psychophysics*, *63*, 361–376.
- Cronbach, L. J. (1942). Studies of acquiescence as a factor in the true-false test. *Journal of Educational Psychology*, *33*, 401–415.
- Erev, I. (1998). Signal detection by human observers: A cutoff reinforcement learning model of categorization decisions under uncertainty. *Psychological Review*, *105*, 280–298.
- Gilbert, D. (1991). How mental systems believe. *The American Psychologist*, *46*, 107–119.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Healy, A. F., & Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 344–354.
- Hodgins, H. S., & Zuckerman, M. (1993). Beyond selecting information: Biases in spontaneous questions and resultant conclusions. *Journal of Experimental Social Psychology*, *29*, 387–407.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.
- Jackson, D. N., & Messick, S. (1958). Content and style in personality assessment. *Psychological Bulletin*, *55*, 243–252.
- Kareev, Y., Arnon, S., & Horwitz-Zeliger, R. (2002). On the misperception of variability. *Journal of Experimental Psychology: General*, *131*, 287–297.
- Kareev, Y., & Fiedler, K. (2006). Nonproportional sampling and the amplification of correlations. *Psychological Science*, *17*, 715–720.

⁵ We thank another anonymous reviewer for raising this point.

- Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis-testing. *Psychological Review*, *94*, 211–228.
- Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*, 379–386.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York: Plenum.
- Kruglanski, A. W. (1990). Lay epistemic theory in cognitive social psychology. *Psychological Inquiry*, *1*, 181–197.
- Kubovy, M., & Healy, A. F. (1977). Decision rule in probabilistic categorization—what is it and how it is learned. *Journal of Experimental Psychology: General*, *106*, 427–446.
- Lopes, L. L. (1982). Doing the impossible: A note on induction and experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*, 626–636.
- Maddox, W. T. (2002). Towards a unified theory of decision criterion learning in perceptual categorization. *Journal of the Experimental Analysis of Behavior*, *78*, 567–595.
- Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1459–1482.
- McKenzie, C. R. M. (2004). Framing effects in inference tasks—and why they are normatively defensible. *Memory & Cognition*, *32*, 874–885.
- McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin & Review*, *7*, 36–366.
- McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*, 33–61.
- McKenzie, C. R. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, *10*, 596–602.
- Messick, S., & Jackson, D. N. (1961). Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, *21*, 771–790.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–631.
- Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review*, *103*, 381–391.
- Selten, R., Abbink, K., & Cox, R. (2005). Learning direction theory and the winner's curse. *Experimental Economics*, *8*, 5–20.
- Selten, R., & Stoecker, R. (1986). End behavior in sequences of finite Prisoner's Dilemma supergames. *Journal of Economic Behavior and Organization*, *7*, 47–70.
- Sher, S., & McKenzie, C. R. M. (2006). Information leakage from logically equivalent frames. *Cognition*, *101*, 467–494.
- Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology*, *43*, 22–34.
- Trope, Y., & Liberman, A. (1996). Social hypothesis-testing: Cognitive and motivational mechanisms. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239–270). New York: Guilford Press.
- Wallsten, T. S., Bender, R. H., & Li, Y. L. (1999). Dissociating judgment from response processes in statement verification: The effects of experience on each component. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 96–115.
- Wallsten, T. S., & González-Vallejo, C. (1994). Statement verification: A stochastic model of judgment and response. *Psychological Review*, *101*, 490–504.
- Zuckerman, M., Knee, C. R., Hodgins, H. S., & Miyake, K. (1995). Hypothesis confirmation: The joint effect of positive test strategy and acquiescence response set. *Journal of Personality and Social Psychology*, *29*, 387–407.