Rational models as theories – not standards – of behavior

Craig R.M. McKenzie

Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0109, USA

When people's behavior in laboratory tasks systematically deviates from a rational model, the implication is that real-world performance could be improved by changing the behavior. However, recent studies suggest that behavioral violations of rational models are at least sometimes the result of strategies that are well adapted to the real world (and not necessarily to the laboratory task). Thus, even if one accepts that certain behavior in the laboratory is irrational, compelling evidence that real-world behavior ought to change accordingly is often lacking. It is suggested here that rational models be seen as theories, and not standards, of behavior.

Cognitive scientists who study reasoning and decision-making often compare people's behavior with a rational model of a given task, and the typical published result is that behavior deviates from the model [1,2]. Such research is popular in part because it implies that it can help people make better inferences and decisions in the real world. In other words, it suggests how behavior ought to change.

However, interpreting behavioral violations of rational models has been a magnet for controversy. For example, it has been argued that participants construe tasks differently from experimenters [3,4] or that experimenters compare behavior with the wrong standard ([5–7]; see [8,9] for reviews). Some researchers have even put forward non-traditional views of rationality to explain deviations from traditional rational models [10,11]. In all of these cases, the implication is that behavior need not change because it is rational after all.

This article briefly reviews recent studies that also question what behavioral violations of rational models mean. However, rather than revising conclusions about the rationality of people's behavior – or about the nature of rationality itself – the findings are used to highlight the fact that, even if one accepts that the violations are irrational, they nonetheless fail to provide compelling evidence that people's real-world behavior ought to change. The reason, in a nutshell, is that errors in the laboratory often appear to be the result of strategies that in fact work well outside the laboratory.

Inference tasks and rarity

Assessing covariation

Consider first research that has examined how people assess whether variables are related, or co-vary. In a typical covariation task, the two variables to be assessed can be either present or absent. For example, participants might be asked to assess the relationship between a medical treatment and recovery from an illness, given the number of people who (a) received the treatment and recovered, (b) received the treatment and did not recover, (c) did not receive the treatment and recovered, and (d) did not receive the treatment and did not recover. The first frequency corresponds to the joint presence of the variables, and the last frequency to their joint absence. Assessing covariation underlies such fundamental behavior as learning, categorization, and judging causation, to name just a few. Traditional rational models of this task are statistical summaries of the four frequencies and consider each frequency equally important [12]. However, decades of research have shown that participants' judgments are influenced most by the number of joint presence observations and are influenced least by the number of joint absence observations [13–15]. This behavior is considered irrational [16,17].

Nonetheless, such behavior appears to make good sense outside the laboratory, where forming and updating beliefs about relationships between variables is arguably more useful than reporting statistical summaries of them. In particular, if (a) participants approach a covariation task as one of (Bayesian) inference, in which they are trying to generalize about the relationship beyond the four frequencies to a larger population of instances, and (b) they assume that, in the larger population, the presence of the variables is rare ($p < 0.5$) and their absence is common ($p > 0.5$), then joint presence observations are in fact more informative than joint absence observations [10,18,19].

To see why rarity matters from a Bayesian viewpoint, imagine that you are trying to determine whether or not there is a relationship between smoking and cancer. The fact that relatively few people smoke and relatively few people have cancer implies that observing a smoker with cancer provides stronger evidence of a relationship than does observing a non-smoker without cancer. This is because a non-smoker without cancer is likely to be observed regardless of whether smoking and cancer are related because there are many nonsmokers and many people without cancer. Such an observation is not much help in determining whether or not there is a relationship. However, although you would be unlikely to observe a smoker with cancer even if there were a relationship, such an observation would be *extremely* unlikely if there were no relationship. In Bayesian terms, the likelihood ratio,

Corresponding author: Craig R.M. McKenzie (cmckenzie@ucsd.edu).

http://tics.trends.com 1364-6613/$ - see front matter © 2003 Elsevier Ltd. All rights reserved. doi:10.1016/S1364-6613(03)00196-7

which captures the impact of evidence, is larger when two rare events occur together than when two common events occur together.

Furthermore, the assumption that the presence of variables is generally rarer than their absence seems reasonable, considering the types of variables people usually think and talk about: just as most people do not smoke and most do not have cancer, it is also the case that most people do not have a fever, most things are not red, and so on [10,18–21]. Compelling evidence in favor of this Bayesian account of covariation assessment is the fact that participants consider joint absence observations more informative than joint presence observations when it is clear that absence of the variables is rare [19]. Other accounts that consider joint presence observations to be normatively most informative for reasons other than rarity cannot explain this latter finding [22].

In short, participants violate the traditional rational model of covariation assessment – and the traditional rational model is the correct one given the typical task instructions to assess a relationship based on only the four cell frequencies – but participants' irrationality is the result of an inferential (Bayesian) strategy that has broader applicability in the real world, suggesting that there is no compelling reason that people ought to change their behavior.

The selection task

Related results have been found for the 'selection task' [23], in which participants test a rule of the form '*If P, then Q*.' To do so, they are shown four cards and must select which cards to turn over to see if the rule is true. For example, imagine testing the rule, '*If there is a vowel on one side, then there is an even number on the other side*.' Each of the four cards has a number on one side and a letter on the other: one card shows an A, one K, one 2, and one 7. Which of those cards must be turned over to see if the rule is true or false? Standard logic dictates (according to one interpretation of the rule) that the A and 7 (the P and not-Q) cards should be turned over because only these potentially reveal the falsifying vowel/odd number combination. Typically, fewer than 10% of participants select only the logically correct cards; instead, they tend to prefer the A and 2 (P and Q) cards (i.e., those mentioned in the rule). This is considered to be a classic demonstration of irrationality.

Again, though, such behavior nonetheless appears to make good sense outside the laboratory, where the world is probabilistic, not deterministic, and has predictable structure. Specifically, the P and Q cards are most informative if participants approach the task as one of (Bayesian) inference, in which (a) they are trying to generalize about the rule beyond the four cards to a larger population of instances, and (b) it is assumed that P and Q are rare relative to not-P and not-Q [20,21,24]. Evidence that it is generally reasonable in the real world to treat P and Q as rare comes from a study showing that people tend to phrase conditional hypotheses in terms of rare, not common, events [25]. Finally, the Bayesian approach predicts that participants will be more likely to turn over the not-Q card when the Q event is more common, and this prediction too has been confirmed ([26–29]; for additional evidence of participants' sensitivity to rarity in inference tasks, see [30,31]).

In short, granting that a joint presence bias in covariation assessment and that turning over the cards mentioned in the rule in the selection task are irrational in the laboratory (when task instructions make clear that only the four cell frequencies or the four cards are of interest), the behavior nonetheless appears to make good sense outside the laboratory. A Bayesian approach, combined with reasonable assumptions about the structure of the natural environment, can explain these errors in the laboratory. Thus, irrational behavior in these tasks does not provide compelling evidence that real-world behavior ought to change.

Framing effects

Framing effects are said to occur when logically equivalent redescriptions of events or objects lead to different preferences or judgments. The best known examples involve choosing between a risky and a riskless option that are described in terms of either gains or losses [32–35], but framing effects also occur with simpler tasks (for reviews, see [36,37]). For example, a medical treatment described as resulting in '75% survival,' rather than the logically equivalent '25% mortality,' will be seen more favorably. Such effects are widely considered to be irrational.

However, recent findings show that speakers do not choose randomly among logically equivalent frames when describing events or objects. Instead, speakers are systematically influenced by background conditions than can be relevant to the judgment or decision. In other words, a speaker's *choice of frame* can convey relevant information [38]. Using the above medical example, it was shown that speakers were more likely to select the '75% survival' frame to describe a new treatment outcome if it led to a higher, rather than a lower, survival rate relative to an old treatment. Generally, speakers are more likely to use a label (e.g. '*X*% survival') when it is above their reference point than when it is below their reference point. To take a more intuitive example, people are more likely to describe a glass as 'half empty' (rather than 'half full') if it used to be full than if it used to be empty (because the glass's emptiness has increased). Thus, different reference-point information can be conveyed through logically equivalent frames, and this information can be relevant: describing the treatment in terms of how many people survive signals that the speaker considers the treatment relatively successful. Furthermore, listeners are sensitive to the information conveyed by a speaker's choice of frame. For example, participants were more likely to infer that an old treatment led to a lower survival rate when the new treatment was described in terms of how many survived than when it was described in terms of how many died [38].

Like the covariation and selection-task accounts, this account of framing effects is essentially Bayesian: if the probability of a speaker choosing a particular frame is greater when certain background conditions hold than when they do not, then the listener can safely infer that the probability that those background conditions hold is greater when that frame is chosen than when it is not. Although this Bayesian account can explain the simpler framing effects, it remains to be seen if it can be fruitfully applied to tasks involving choices between risky and riskless options.

Also consistent with the previous analyses, the current analysis implies that framing effects do not imply that real-world behavior ought to change. Ignoring the information conveyed by speakers' choice of frame would put listeners at a disadvantage under ordinary circumstances. However, unlike the covariation and selection task analyses, the current one implies that framing effects might not be irrational, even in the laboratory. Nonetheless, common to all the analyses is the application of a different rational (Bayesian) model *and* taking into account real-world conditions in order to understand 'errors' in the laboratory.

Rational rules and real-world accuracy

So far I have argued that irrational behavior in the laboratory does not necessarily imply that real-world behavior ought to change. Now I want to speculate about the following possibility: irrational behavior might never imply poor real-world performance (see also [39]).

It is generally accepted that behaving rationally – that is, following rational rules – and being accurate in the real world are not the same thing (e.g. [40–43]). In fact, although there is a large amount of evidence showing that people violate rational rules [1,2], research examining real-world judgments has often concluded that such judgments are surprisingly accurate (e.g. [40,44,45]), even though they are often based on very little information and the judges have little or no insight into how they made them [44].

Could it be that following rules is not the key to real-world accuracy? Of interest is the observation that research on artificial intelligence (AI), which implements rules (although not necessarily rational rules) in the form of computer programs in an attempt to perform real-world tasks, has been plagued by failure [46]. Despite early claims that machines would be able to rival – even exceed – human performance, this has not turned out to be the case, except in highly constrained, well-defined environments, such as playing chess. Interestingly, the benchmark in AI is human behavior – and this benchmark is virtually never reached. Given that computers are 'logic machines,' it is interesting that it is so difficult to get them to do tasks that we routinely perform, such as understand a story, produce and understand speech, and recognize scenes. Functioning in the real world requires 'common sense', which might be impossible, in principle, to capture in rules [46].

Thus, not only might following rules – rational or otherwise – fail to guarantee real-world accuracy, the two might not even be compatible. In fact, scholars in diverse fields have reached a similar conclusion: depending on a purely logical analysis will not get you very far in the real world, where context, meaning and relevance, rather than pure structure, are crucial [46–48]. It is generally agreed that the cognitive system's most fascinating quality is its ability to solve apparently intractable problems with such apparent ease [49]. How it does so largely remains a mystery, but the failings of AI suggest that following rules is not the key. To the extent that following rational rules does not entail real-world accuracy, evidence that people fail to follow these rules is not evidence that real-world behavior ought to change.

This leaves us, however, with an apparent paradox: if being Bayesian requires following rules, and following rules is not the key to real-world accuracy, then how can Bayesianism lead to real-world accuracy (as I argued earlier)? First, I am not saying that people are optimal (rule-following) Bayesians [14]. Second, Bayesian models are notorious for their enormous complexity when applied to even modestly complicated real-world problems, suggesting that there can be no optimal Bayesians in the real world. My Bayesian account is purely qualitative (and hence feasible) and allows (even assumes) sensitivity to context. In the case of covariation assessment and the selection task, I claimed that (a) people are sensitive to, among other things, the rarity of data, (b) they make assumptions about rarity that generally reflect the natural environment, and (c) these assumptions are overridden when it is clear that they do not apply. Normative principles, combined with environmental considerations, provide good *guides* for understanding behavior.

Conclusion

At least some behavioral violations of rational models in the laboratory appear to result from behavior that is well-suited to the natural environment, implying that evidence of irrationality is not sufficient for concluding that real-world behavior ought to change. Furthermore, it is possible that following rational rules is not even the key to accuracy in the real world, suggesting that irrational behavior might never indicate that real-world behavior ought to change. These observations lead to a natural prescription for researchers: treat rational models as theories, not as standards, of behavior. Viewing rational models as standards implies that behavior ought to change when they are violated, whereas viewing them as theories does not. It also opens up the possibility of testing multiple rational models in order to see which provides the best account of behavior, rather than comparing behavior with a single model [50]. In each empirical example discussed earlier, a rational model that was not considered until recently can explain the otherwise puzzling behavior. Rational models serve as useful guides for understanding behavior, especially when they are combined with considerations of the environment. When a rational model fails to describe behavior, a different rational model, not different behavior, might be called for.

Acknowledgements

Preparation of this article was supported by NSF grants SES-0079615 and SES-0242049. The author gratefully acknowledges helpful comments on earlier drafts from Mike Liersch, Shlomi Sher, and the anonymous reviewers.

References

1 Gilovich, T., *et al.* eds (2002) *Heuristics and biases: The Psychology of Intuitive Judgment*, Cambridge University Press

2 Kahneman, D. and Tversky, A. eds (2000) *Choices, Values, and Frames*, Cambridge University Press

3 Hilton, D.J. (1995) The social context of reasoning: conversational inference and rational judgment. *Psychol. Bull.* 118, 248–271

4 Schwarz, N. (1996) *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*, Erlbaum

5 Birnbaum, M.H. (1983) Base rates in Bayesian inference: signal detection analysis of the cab problem. *Am. J. Psychol.* 96, 85–94

6 Gigerenzer, G. (1991) How to make cognitive illusions disappear: beyond 'heuristics and biases'. *Eur. Rev. Social Psychol.* 2, 83–115

7 Lopes, L.L. (1982) On doing the impossible: a note on induction and the experience of randomness. *J. Exp. Psychol. Learn. Mem. Cogn.* 8, 626–636

8 Stanovich, K.E. (1999) *Who is Rational? Studies of Individual Differences in Reasoning*, Erlbaum

9 Stanovich, K.E. and West, R.F. (2000) Individual differences in reasoning: implications for the rationality debate? *Behav. Brain Sci.* 23, 645–726

10 Anderson, J.R. (1990) *The Adaptive Character of Thought*, Erlbaum

11 Chase, V.M. *et al.* (1998) Visions of rationality. *Trends Cogn. Sci.* 2, 206–214

12 Allan, L.G. (1980) A note on measurement of contingency between two binary variables in judgment tasks. *Bull. Psychon. Soc.* 15, 147–149

13 Levin, I.P. *et al.* (1993) Multiple methods for examining biased information use in contingency judgments. *Organ. Behav. Hum. Decis. Process.* 55, 228–250

14 McKenzie, C.R.M. (1994) The accuracy of intuitive judgment strategies: covariation assessment and Bayesian inference. *Cogn. Psychol.* 26, 209–239

15 Wasserman, E.A. *et al.* (1990) Contributions of specific cell information to judgments of interevent contingency. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 509–521

16 Kao, S-F. and Wasserman, E.A. (1993) Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 1363–1386

17 Mandel, D.R. and Lehman, D.R. (1998) Integration of contingency information in judgments of cause, covariation, and probability. *J. Exp. Psychol. Gen.* 127, 269–285

18 Anderson, J.R. and Sheu, C-F. (1995) Causal inferences as perceptual judgments. *Mem. Cogn.* 23, 510–524

19 McKenzie, C.R.M. and Mikkelsen, L.A. A Bayesian view of covariation assessment. *Cogn. Psychol.* (in press)

20 Oaksford, M. and Chater, N. (1994) A rational analysis of the selection task as optimal data selection. *Psychol. Rev.* 101, 608–631

21 Oaksford, M. and Chater, N. (1996) Rational explanation of the selection task. *Psychol. Rev.* 103, 381–391

22 Cheng, P.W. (1997) From covariation to causation: a causal power theory. *Psychol. Rev.* 104, 367–405

23 Wason, P.C. (1966) Reasoning. In *New Horizons in Psychology* (Foss, B.M., ed.), pp. 135–161, Penguin

24 Nickerson, R.S. (1996) Hempel's paradox and Wason's selection task: logical and psychological puzzles of confirmation. *Think. Reason.* 2, 1–31

25 McKenzie, C.R.M. *et al.* (2001) Do conditional hypotheses target rare events? *Organ. Behav. Hum. Decis. Process.* 85, 291–309

26 Oaksford, M. and Chater, N. (2003) Optimal data selection: revision, review, and re-evaluation. *Psychon. Bull. Rev.* 10, 289–318

27 Oaksford, M. *et al.* (1999) Probabilistic effects in data selection. *Think. Reason.* 5, 193–243

28 Oaksford, M. *et al.* (1997) Optimal data selection in the reduced array selection task (RAST). *J. Exp. Psychol. Learn. Mem. Cogn.* 23, 441–458

29 Oaksford, M. and Wakefield, M. (2003) Data selection and natural sampling: probabilities do matter. *Mem. Cogn.* 31, 143–154

30 McKenzie, C.R.M. and Amin, M.B. (2002) When wrong predictions provide more support than right ones. *Psychon. Bull. Rev.* 9, 821–828

31 McKenzie, C.R.M. and Mikkelsen, L.A. (2000) The psychological side of Hempel's paradox of confirmation. *Psychon. Bull. Rev.* 7, 360–366

32 Kahneman, D. and Tversky, A. (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–291

33 Kahneman, D. and Tversky, A. (1984) Choices, values, and frames. *Am. Psychol.* 39, 341–350

34 Tversky, A. and Kahneman, D. (1981) The framing of decisions and the psychology of choice. *Science* 211, 453–458

35 Tversky, A. and Kahneman, D. (1986) Rational choice and the framing of decisions. *J. Bus.* 59, S251–S278

36 Kühberger, A. (1998) The influence of framing on risky decisions: a meta-analysis. *Organ. Behav. Hum. Decis. Process.* 75, 23–55

37 Levin, I.P. *et al.* (1998) All frames are not created equal: a typology and critical analysis of framing effects. *Organ. Behav. Hum. Decis. Process.* 76, 149–188

38 McKenzie, C.R.M. and Nelson, J.D. What a speaker's choice of frame reveals: reference points, frame selection, and framing effects. *Psychon. Bull. Rev.* (in press)

39 McKenzie, C.R.M. Judgment and decision making. In *Handbook of Cognition* (Lamberts, K. and Goldstone, R.L., eds.), Sage (in press)

40 Funder, D.C. (1987) Errors and mistakes: evaluating the accuracy of social judgment. *Psychol. Bull.* 101, 75–90

41 Gigerenzer, G. and Goldstein, D.G. (1996) Reasoning the fast and frugal way: models of bounded rationality. *Psychol. Rev.* 103, 650–669

42 Gigerenzer, G., Todd, P.M. and the ABC Research Group (1999) *Simple Heuristics that Make Us Smart*, Oxford University Press

43 Hammond, K.R. (1996) *Human Judgment and Social Policy*, Oxford University Press

44 Ambady, N. *et al.* (2000) Toward a histology of social judgment behavior: judgmental accuracy from thin slices of the behavioral stream. In *Advances in Experimental Social Psycholog* (Zanna, M.P., ed.), pp. 201–271, Academic Press

45 Wright, J.C. and Drinkwater, M. (1997) Rationality vs. accuracy of social judgment. *Social Cogn.* 15, 245–273

46 Dreyfus, H.L. (1992) *What Computers Still Can't Do: A Critique of Artificial Reason*, MIT Press

47 Damasio, A.R. (1994) *Descartes' Error: Emotion, Reason, and the Human Brain*, Avon

48 Devlin, K. (1997) *Goodbye, Descartes: The End of Logic and the Search for a New Cosmology of the Mind*, Wiley

49 Medin, D.L. *et al.* (2001) *Cognitive Psychology*, 3rd edn, Harcourt

50 Hertwig, R. and Todd, P.M. (2000) Biases to the left, fallacies to the right: stuck in the middle with null hypothesis significance testing. *Psycoloquy* 11 (28)