Cognitive Science 30 (2006) 1081–1095

Decision Making and Confidence Given Uncertain Advice

Michael D. Lee[a], Matthew J. Dry[b]

[a]*Department of Cognitive Sciences, University of California, Irvine*
[b]*School of Psychology, University of Adelaide*

Received 21 November 2005; received in revised form 3 March 2006; accepted 8 March 2006

Abstract

We study human decision making in a simple forced-choice task that manipulates the frequency and accuracy of available information. Empirically, we find that people make decisions consistent with the advice provided, but that their subjective confidence in their decisions shows 2 interesting properties. First, people's confidence does not depend solely on the accuracy of the advice. Rather, confidence seems to be influenced by both the frequency and accuracy of the advice. Second, people are less confident in their guessed decisions when they have to make relatively more of them. Theoretically, we develop and evaluate a type of sequential sampling process model—known as a self-regulating accumulator—that accounts for both decision making and confidence. The model captures the regularities in people's behavior with interpretable parameter values, and we show its ability to fit the data is not due to excessive model complexity. Using the model, we draw conclusions about some properties of human reasoning under uncertainty.

Keywords: Decision making; Uncertainty; Confidence; Accumulator model; Sequential sampling process

1. Introduction

Most decisions in the real world must be made under conditions of uncertainty, and so understanding how people reason with incomplete and inaccurate information is a central problem for cognitive psychology. One major line of research is concerned with the optimality of human decisions under uncertainty with respect to what Simon (1976) termed *substantively rational* inference, from the perspectives of both normative decision theory, and, more recently, Bayesian methods (e.g., Kahneman & Tversky, 2000; Nickerson, 2004; Tenenbaum & Griffiths, 2001). A complementary line of research has developed and evaluated what Simon (1976) termed *procedurally rational* inference, providing accounts of heuristic processes that make fast and accurate decisions based on uncertain information (e.g., Gigerenzer & Selten, 2001; Gigerenzer & Todd, 1999).

Correspondence should be addressed to Michael D. Lee, Department of Cognitive Sciences, University of California, Irvine, CA 92697–5100. E-mail: mdlee@uci.edu

In this article, we focus on the interaction and relative contributions of two fundamental sources of uncertainty that pervade most human decision-making situations. These are the uncertainty that arises because of incomplete information, and the uncertainty that arises because of inaccurate information. The task we consider is one in which people must make a sequence of two-choice, forced-choice decisions, on the basis of uncertain advice. On each trial, people must choose between a left door and a right door, one of which contains treasure. The information people are given on the trial either takes the form of prescriptive advice (being told to "go left" or "go right"), or the absence or advice (being told the advisor has "no idea"). We manipulate the frequency and accuracy of the advice provided in realistic ways, and assess both the decisions people make and the confidence they have in their decisions over a series of trials.

A reasonable set of expectations for the decisions people will make in this task might involve two obvious predictions. The first is to expect people to follow advice when it is offered, as long as the advice has proved better than chance in the past. The second is to expect people to guess when no advice is provided, provided there is no base rate information that favors one choice over the other.

Forming expectations for the confidence that people will have in their decisions is less straightforward. Perhaps the most obvious prediction involves equating confidence with accuracy. This idea dates back in psychology to at least the verbal model of Peirce and Jastrow (1884), and is explicitly adopted in some prominent contemporary models, such as the probabilistic mental models of Gigerenzer, Hoffrage, and Kleinbölting (1991). Under this conception, both choice and confidence behaviors are determined by the conditional probability that a decision is correct given the available information. Procedurally, a cue validity is associated with information such as "go left" and "go right" advice, being defined as the proportion of times that information has proven correct in the past. The models then assume that people choose the highest validity cue, and their confidence is an expression of the validity value itself. Gigerenzer et al. (1991) argued that these assumptions are both "rational and simple" (p. 509).

This cue validity account makes clear predictions about confidence for our task. It predicts, on those trials where people accept advice, they will become more confident as the accuracy of advice improves. It also predicts that people will have the same (low) confidence in guessed decisions on trials where no advice is offered, regardless of how many of these trials are encountered. We use these plausible predictions to make it clear why the results of our experiment are interesting, and to motivate the importance of the model we develop to account for the data.

2. Experiment

2.1. Participants

A total of 30 participants (14 men, 16 women), with a mean age of 25 years completed the experiment.

2.2. Procedure

On each trial, the participant was presented with a graphic display of two doors—a left door and a right door—and was told there was treasure behind one of the doors, and nothing behind the other. Participants received advice, in the form of a text message above the doors, that read "go left," "go right," or "no idea." Participants then decided which door contained the treasure, and rated their confidence in their decision on a 5-point scale ranging from 1 (*guess*) to 5 (*sure*). Once the decision and confidence rating had been made, the correct answer was revealed graphically, and the next trial commenced.

A within-participants design was used, with each participant completing 50 decision-making trials for each of six conditions. The six conditions were completed in random order, and differed systematically in the accuracy and frequency of advice provided. On each trial, the advisor was given probabilistic knowledge of the correct location of the treasure. For example, the advisor might know that there was a 85% probability that the treasure was behind the right door. This probabilistic advice was determined by drawing independently from the uniform distribution on [0,1] for each trial. The true location of the treasure was determined by a second independent draw from the uniform distribution on [0,1] for each trial being compared with the original probabilistic advice value. This means that if the advisor had probabilistic knowledge that there was a 85% probability that the treasure was behind the right door, there was indeed an 85% probability that the treasure would actually be behind that door for that trial.

The different conditions manipulated how certain the advisor's probabilistic knowledge had to be before "go left" or "go right" advice was given, rather than "no idea" being produced. In the most extreme condition, advice was offered on every trial, according to which door had the greater probability. That is, whichever choice had greater than a 50% chance was offered. In the next condition, "go left" or "go right" advice was only offered if it was at least 60% likely this was correct advice, otherwise "no idea" was produced. Successive conditions became more conservative in this cut-off, requiring 70%, 80%, and 90% certainty before offering advice. In the final condition, conservatism was taken to the extreme, and "no idea" was always produced. This final condition corresponds to requiring an unattainable 100% certainty before offering advice.

It is worth noting that a natural consequence of this design is that the left and right alternatives are symmetric, and so (on average) advice is offered relating to each door equally often in each condition. Accordingly, participants were told that the location of the treasure was, over all trials, equally likely to be behind either door, and the correct answer was determined independently on each trial. The independence of the six conditions was also emphasized to participants, with an enforced break, a verbal reminder, and the use of different background colors in each condition.

2.3. Results

Fig. 1 shows the basic patterns of empirical confidence and decision-making behavior, averaged across all participants, for each of the six experimental conditions. Because of the

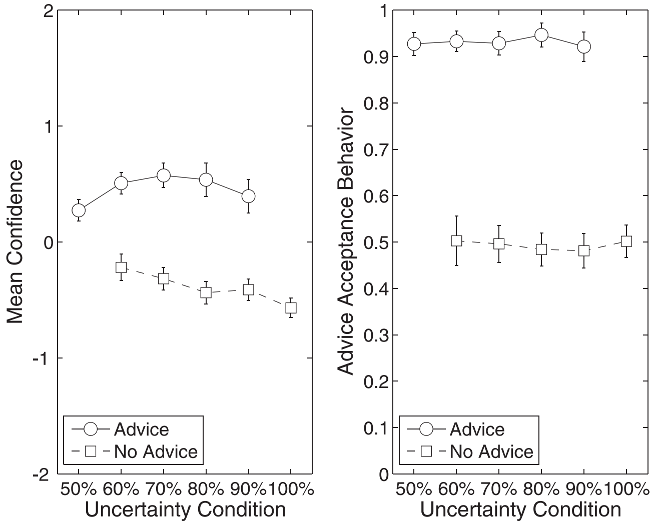

Fig. 1. The experimental data. The left panel shows the pattern of change of mean confidence for advice trials and no advice trials across the six experimental conditions. The right panel shows the pattern of change in advice acceptance behavior across the six conditions, showing the proportion of advice accepted for advice trials, and the proportion of "go left" decisions for no advice trials. Error bars show one standard error.

within-participants design, we consider normalized confidence measures, calculated as the z score for each participant.[1] These z scores are then averaged across participants to give an overall mean for each condition.[2] The left panel shows the pattern of change of mean confidence for advice trials and no advice trials. The right panel shows the pattern of change in advice acceptance behavior. For advice trials, the mean proportion of trials for which the advice was accepted is shown. For no advice trials, the mean proportion of "go left" decisions is shown.

In three respects, these data agree with the expectations we outlined earlier. Overwhelmingly, people accept advice when it is offered, make decisions consistent with guessing when no advice is offered, and are more confident in their decision in the first case when compared to the second.

There are, however, two surprising regularities in the data. The first is that mean confidence on no advice trials decreases across the six conditions. This shows that people become less confident in their guessed decisions when they have to make them relatively often. This is not consistent with equating confidence and accuracy, because the guesses are (on average) equally accurate across all of the conditions.

The second surprise is that mean confidence on advice trials does not increase across the six conditions. It is relatively constant, and perhaps even shows a slight inverted U-shape. The lack of increase shows that people do not become more confident in those decisions based on advice as the accuracy of that advice improves. This finding directly contradicts the prediction of the cue validity account, which argued that people will become more confident in their decisions as the validity of advice increases. A simple calculation shows that the conditional probability cue validities are .75, .80, .85, .90, and .95 for the five conditions where advice is provided, starting at 50% and increasing up to 90%. The left panel of Fig. 1 shows, however, that, from a high point in the 70% condition, people become less confident in their decisions as both the accuracy (moving toward the 50% condition) or the frequency (moving toward the 90% condition) of the advice deteriorates.

3. An Accumulator Sequential Sampling Model

Sequential sampling models are long established and widely used process accounts of human decision making for a variety of tasks, ranging from simple perceptual judgment tasks, to more abstract cognitive tasks involving choice and classification (e.g., Busemeyer & Rapoport, 1988; Busemeyer & Townsend, 1993; Diederich, 1997; Laming, 1968; Lee & Corlett, 2003; Lee & Cummins, 2004; Link & Heath, 1975; Nosofsky & Palmeri, 1997; Ratcliff, 1978; Smith, 2000; Vickers, 1979; Wallsten & Barton, 1982).

Within the general sequential sampling approach, we use the self-regulating accumulator model developed by Vickers (1979), and summarized by Vickers and Lee (1998). Given the prevalence and popularity of alternative random-walk and diffusion sequential sampling models, it is probably worth justifying and motivating our use of the accumulator approach. First, the accumulator model provides a principled approach to modeling subjective confidence, through its balance of evidence mechanism, that has been demonstrated to be successful empirically (Baranski & Petrusic, 1994; Smith & Vickers, 1988; van Zandt, 2000; Vickers & Packer, 1982; Vickers, Smith, Burt, & Brown, 1985). Second, we wish to model the way people adapt their decision making over repeated trials, and so need an account of the way people change in response to learning about the environment, and self-regulate their decision processes. An attractive property of the self-regulating accumulator model is that it incorporates an account of the adaptation process within the model itself, rather than relying on less constrained post-hoc parametric variation to explain these changes.

3.1. Evidence distributions

Accumulator models, like all sequential sampling models, assume that indivduals make decisions by sampling evidence for the left and right alternatives, based on their current knowledge of the task environment. This involves remembering separate probability distributions representing the evidence provided by "go left," "go right," and "no idea" advice. We assume that initially participants are maximally ignorant about the rate that "go left" or "go right" advice will be correct, but that the "no idea" advice conveys the information that either choice could possibly be correct.

We then assume the probability distributions associated with "go left" and "go right" advice are modified as this advice proves to be good (i.e., correct) or bad (i.e., incorrect). This information is summarized by four counts: g_l, the number of times "go left" advice has proven to be good; b_l, the number of times "go left" advice has proven to be bad; g_r, the number of times "go right" advice has proven to be good; and b_r, the number of times "go right" advice has proven to be bad. Using these counts, the probability distributions for "go left" and "go right" advice are updated using Bayes theorem.[3] For the "no idea" advice, however, we assume that no updating takes place.

3.2. Sequential sampling process

Given the evidence distribution corresponding to the presented advice, the accumulator sequential sampling process proceeds according to the flow diagram shown in Fig. 2. At the beginning of a series of trials (i.e., at the beginning of a condition), criterion evidence totals, k_l and k_r, for making left and right decisions, respectively, are both set to a value κ.

Advice is then provided, corresponding to a probability distribution. On each iteration, an independent sample is taken from this evidence distribution on the log-odds scale, as the model adds successive evidence values. On each iteration, if the evidence is positive, it is added to a right accumulator, t_r. If the evidence is negative, its absolute value is added to a left accumulator, t_l. Sampling continues until either the total t_l exceeds the criterion total k_l, or the total t_r exceeds the criterion total k_r.

At this point, the model makes a left or right decision accordingly, with a response time corresponding to the number of iterations. The balance-of-evidence measure of confidence, c, is provided by the difference between the totals, expressed as a proportion of the total evidence accumulated, so that

$$c = |t_l - t_r| \, / \, (t_l + t_r) \, .$$

A graphical representation of the model making a decision on a single trial is shown in Fig. 3. The top panel shows the evidence distribution in memory corresponding to "go right" advice. The operation of the decision-making process is shown in the bottom panel, with the two accumulators shown as evolving solid lines, and the criterion threshold levels of evidence as dotted lines. In this example, it takes a little over 30 time units for a right decision to be made, with relatively high confidence, because most of the accumulated evidence at that point is in favor of a right decision.

3.3. Adaptation

Having made a decision, the model adapts to its environment in two ways. First, if advice was given, the feedback about the correctness of the advice is used to update the relevant memory counts. In other words, feedback is provided about whether or not the decision was correct, allowing one of the g_l, b_l, g_r, and b_r counts to be updated.

Second, the decision confidence is used to self-regulate the criterion threshold levels. The difference between the confidence with which the decision was made, c and a target level of confidence τ is calculated as $h = c - \tau$. If h is positive, it is added to an overconfidence accumulator t_{ol} for left decisions and t_{or} for right decisions. If h is negative, its absolute value is added to an underconfidence accumulator t_{ul} for left decisions and t_{ur} for right decisions.

If any of these over- or underconfidence accumulators exceeds a critical amount γ, the model undertakes a self-regulating adjustment of a decision threshold. This is done by increasing a decision threshold, making underconfident decisions, or decreasing a decision threshold, leading to overconfident decisions, according to a learning rate $0 \leq \lambda \leq 1$ and the difference be-

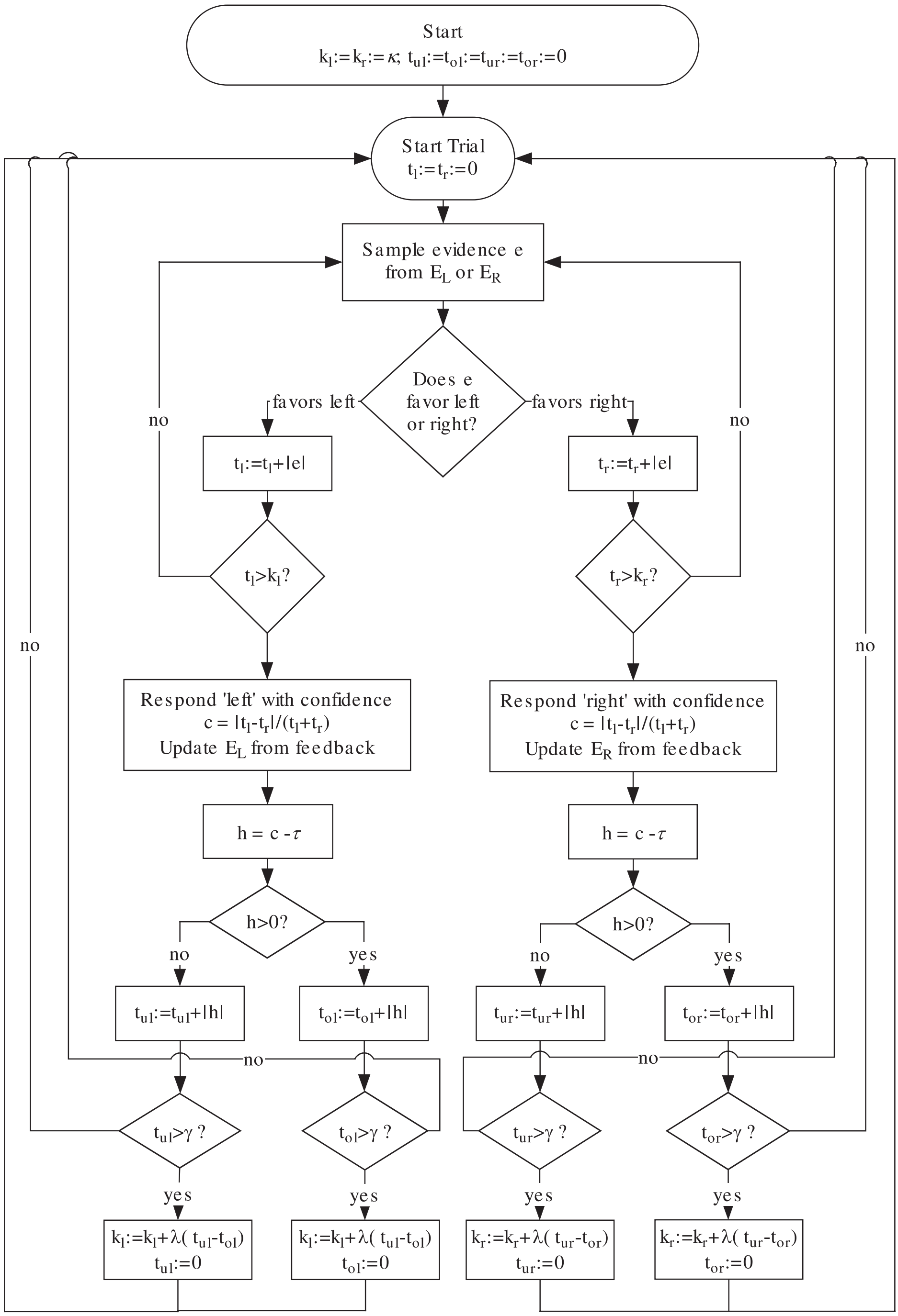

Fig. 2. A flow chart describing the self-regulating accumulator model.

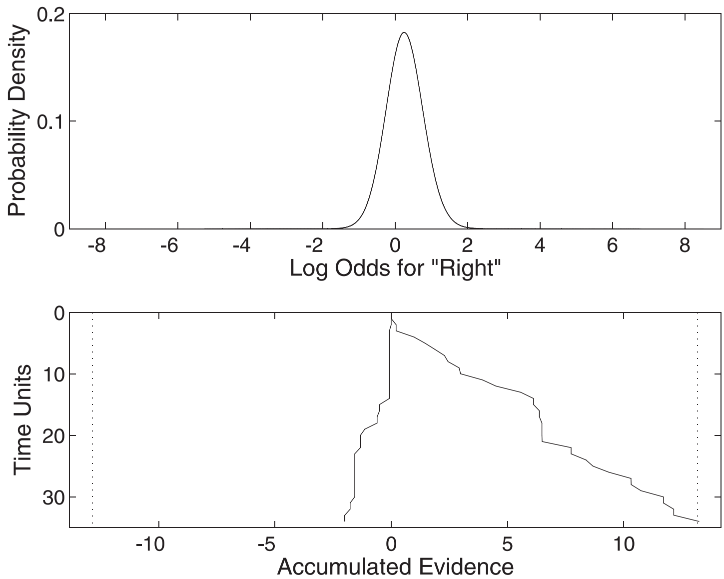

Fig. 3. An accumulator sequential sampling process account of decision making on a single trial.

tween accumulated over- and underconfidence totals. For example, if the underconfidence accumulator for left decisions t_{ul} exceeds the critical amount, then the threshold for making left decisions, t_l is increased by λ $(t_{ul} - t_{ol})$. The other possibilities for adjustment are formulated similarly, and are detailed in Fig. 2.

4. Model evaluation

We evaluate the model in two stages. In the first, we focus on descriptive adequacy, by considering the ability of the model to fit the data at an appropriate parameterization. In the second, we focus on model complexity (e.g., Myung, Forster, & Browne, 2000; Pitt, Myung, & Zhang, 2002; Roberts & Pashler, 2000), and show that the model is highly constrained in the behavior it can produce at reasonable parameterizations.

4.1. Model fitting

The model has four free parameters: κ, the initial evidence level required to make a decision; τ, the target level of confidence; γ, the critical level of over- or underconfidence; and λ, the learning rate.

It is possible to interpret the scale for each of these parameters, and so constrain their values meaningfully. The target level of confidence τ is a value between 0 and 1, which should be set at a value above .5 in a two-choice task. The learning rate λ also lies between 0 and 1, representing the usual trade-off between speed and stability of adaptation. The level of evidence needed to make a decision κ lies on a log-odds scale, and so can be treated in the same way as Bayes factors and other likelihood ratios, which are commonly interpreted on this scale (e.g., Kass & Raftery, 1995). The critical level of over- or underconfidence γ simply accumulates differences on the confidence scale.

Using these interpretations as a guide, we examined the behavior of the model for every possible combination of the following parameter values: $\kappa = \{0, 2, 6, 10\}$, $\tau = \{0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, $\gamma = \{1, 3, 5\}$, and $\lambda = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. At each of these parameterizations, we measured the fit of the advice acceptance behavior and mean confidence of the model to the empirical data.

Fit was measured on a log-likelihood scale using Gaussian likelihood functions with the means and standard deviations shown in Fig. 1, giving equal weight to both the confidence and advice acceptance behavioral measures. Because of the different scales on which empirical confidence and the confidence of the model are measured, the single scalar multiple that best mapped the [0, 1] range of model confidence values onto the [1, 5] range for the empirical data was found in each case.

Fig. 4 shows the best fitting behavior of the model to the data, achieved using the parameterization $\kappa = 2$, $\tau = 1.0$, $\gamma = 3$, and $\lambda = 0.5$. It is clear that the model is able to emulate closely the empirical regularities in which we are interested. As a comparison, the left panel also shows the best fitting confidence behavior of the plausible cue validity model canvassed in the Introduction, according to the same fitting procedure (i.e., using the best possible single scalar to transform validities to confidence measures). This makes clear the incompatibilities of our data with the assumption that confidence increases with advice accuracy and is insensi-

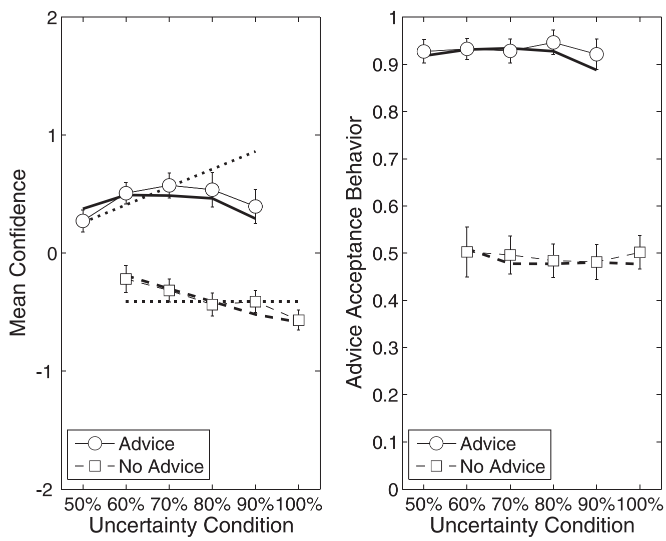

Fig. 4. The fit of the accumulator model (heavy lines) and cue validity model (dotted lines) to the data at the best fitting parameterization. The left panel shows the best fit to mean confidence. The right panel shows the best fit to advice acceptance behavior.

tive to frequency. It also highlights, we believe, why our data are interesting, and why the ability of the accumulator model to fit the data is significant.

4.2. Complexity analysis

One possible explanation for the ability of the accumulator model to fit the data is that it is a very complicated model, potentially able to fit all sorts of qualitatively different patterns of behavior by using different parameterizations. To consider this issue, we examined the data patterns generated by all of the $4 \times 6 \times 3 \times 5 = 360$ parameterizations considered.

Fig. 5 shows the full range of model behavior. For advice trials, there are two possible patterns of change in mean confidence. Both have the slight inverted U-shape of the empirical data, but one is more confident than the other, and shows less change in confidence across conditions.

For no advice trials, there are three distinguishable levels of confidence, one of which is too high to agree with the empirical data. Within the lower bands, some of the model behavior shows a linear decrease in mean confidence across conditions that agrees with empirical data, whereas other behavior shows little or no change in mean confidence.

For decisions, the model is essentially only able to show one pattern of behavior, which involves accepting advice with high probability, and guessing when no advice is provided.

This analysis makes it clear that the model is not complicated, in the formal sense explained by Myung, Balasubramanian, and Pitt (2000), as it is able to index a only small number of distinguishable predictions about confidence and decision-making behavior on the task.

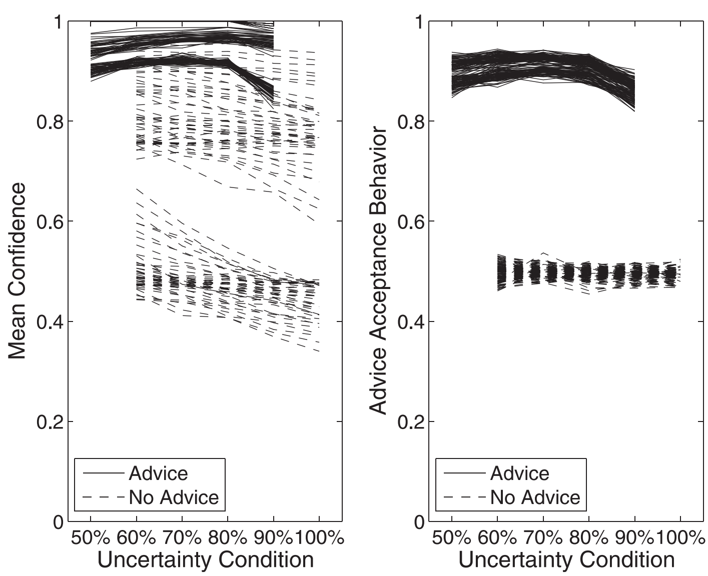

Fig. 5. The complexity of the model, showing its behavior over the set of parameterizations considered, for both advice and no advice trials. The left panel shows confidence behavior. The right panel shows advice acceptance behavior.

5. What the model tells us about human decision making

Our evaluation shows that the model can account for the data, and does so without excessive complexity. What it does not offer is an explanation of *why* the model behaves as it does, and hence why people behave as they do in this task.

Clear insights along these lines are provided by considering which parameter values are responsible for the observed variation in model behavior. If the initial level of evidence to make a decision, κ, is set to the lowest values of 0 or 2, there is high and relatively constant confidence for both advice trials and no advice trials (although, on average, no advice trials are a little lower). This corresponds to the upper bands for both advice and no advice trials in the left panel of Fig. 5. Intuitively, this is because often only one or two samples will be sufficient to make a decision, and no evidence for the alternative choice will be accumulated, leading to perfect confidence. If κ is set to the more conservative levels of 6 and 10, corresponding to requiring positive or strong evidence before making a decision, then the change in mean confidence across conditions falls into the band that agrees with the empirical data, and confidence for no advice trials is low.

For the model to agree with the empirical data on no advice trials, it must show the further pattern of declining as more "no idea" advice is provided. This is achieved by having either a low criterion γ on the critical threshold for adapting to over- or underconfidence, or a high learning rate λ. A decrease in confidence on no advice trials is evident when either of these parameter settings is used, and is most exaggerated when they are both present.

Taken together, these conclusions mean that the model captures the empirical data, as long as (a) a conservative initial decision threshold is used, and (b) the learning rate and threshold for adaptivity allow large changes when self-regulating the decision threshold. These conclusions seem well aligned to the nature of our task, which gave people little relevant prior information, and so forced them to be cautious initially, but then provided them with highly informative feedback, and so encouraged them to adapt quickly.

Indeed, it is possible to provide a meaningful psychological account of both the surprising features of our data by considering the two ways in which the model adapts: through external adaptation to the environment based on feedback, and through internal self-regulation based on confidence.

5.1. External adaptation

Because the evidence in memory almost always leads the model to accept advice, what determines confidence is the extent to which evidence for the alternative choice is sampled. For a trial where "go right" advice is given, this is measured by the extent to which the evidence distribution gives probability to negative log-odds. Equivalently, it depends on the extent to which the probability distribution gives density to probabilities less than .5.

There are two ways in which the evidence distributions can give density to values less than .5. One way is to have a mean near .5, and some variance. This is what happens for the experimental conditions where advice is always offered, and so the counts of correct and incorrect advice are relatively close. The alternative is to have a large variance, whatever the mean. This is what happens for the experimental conditions where little advice is given, and so the counts of correct and incorrect advice are small.

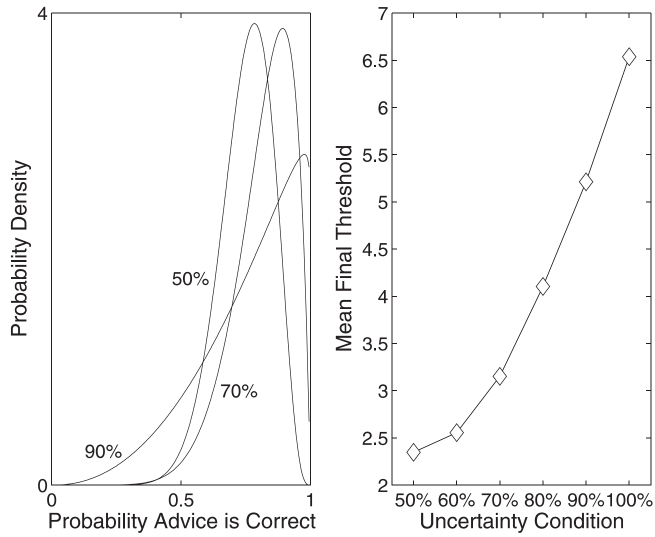

Fig. 6. Adaptive behavior of the accumulator model. The left panel shows the mean evidence distributions after 30 trials for three of the experimental conditions. The right panel shows the change in the mean final decision thresholds across the six experimental conditions, at the best fitting model parameterization.

This state of affairs is represented graphically in the left panel of Fig. 6, which shows the expected probability distributions for the 50%, 70%, and 90% experimental conditions at trial 30 out of 50, under the best fitting parameterization. The 70% distribution gives the least probability to values less than .5 because it is centered at a value far away (due to accurate advice) and has relatively small variance (due to frequent advice).

Intuitively, when the advice is poor, the model considers the alternative, and loses confidence. When little advice is given, the model is unsure, and loses confidence. When a reasonable amount of reasonably accurate advice is given, the model is most confident. This adaptation to the accuracy and frequency of advice is responsible for our model predicting an inverted U-shape in confidence on advice trials.

5.2. Internal adaptation

The right panel of Fig. 6 shows the change in the mean final decision thresholds for each condition, at the best fitting parameterization. As conditions include more no advice trials, and more guessing decisions are made with low confidence, the thresholds are adapted to larger values. This self-regulation is responsible for the decline in confidence for no advice trials.

6. Conclusion

Our central finding is that people are sensitive to the accuracy and frequency of information in assessing the quality of their decisions. This is consistent with some relevant applied research that has argued decision makers are less likely either to trust or follow advice from conservative advisors (Sniezek & van Swol, 2001), and that cautious advisors are rated as less accurate than overconfident advisors, regardless of the actual accuracy of the advice (Price & Stone, 2004). The accumulator model account we have developed and evaluated provides a rational and interpretable explanation of these and our findings.

Given the success of the accumulator model, the obvious line for future research is to consider its ability to account for human performance in a wider range of tasks. In particular, it is important to consider relations between accuracy and frequency under environmental assumptions that differ from those made here. The assumptions we made correspond naturally to one important real-world scenario, in which the information available trades off accuracy with frequency. At one extreme is the case where people are always given prescriptive advice, but this advice is not very accurate. At the other extreme is the case where people are almost never given prescriptive advice, but it is almost always accurate when provided.

Although we believe that, in many environments, sources that provide information less frequently are often the most accurate, this is not universally true. It would be straightforward to consider the same task and model as presented here in environments where accuracy and frequency of information varied in more flexible ways. We hope that by developing and evaluating our model of people's behavior under additional plausible combinations of advice accuracy and frequency, we will provide some further insight into the mechanisms and the basis for the way people reason under uncertainty.

Notes

1. The nonnormalized raw data are extremely similar to the normalized data reported here, and support the same modeling analysis.
2. We examined the possibility of order effects by calculating the difference between the overall means for both decision and confidence measures and the means for these measures in each position in the experimental sequence. This analysis suggested the order in which conditions were completed did not interact systematically with the behavior of the participants.
3. Technically, we rely on established Bayesian results (Jaynes, 2003, pp. 382–385) to set a proper prior that approximates the Haldane prior (i.e., we use Beta $(\varepsilon, \varepsilon)$ with a very small $\varepsilon > 0$) for the "go left" and "go right" advice distributions, and a uniform prior Beta (1, 1) for the "no idea" distribution. Because these are both beta distributions, and are conjugate to the binomial likelihood function, the posteriors for "go left" and "go right" advice are approximately Beta (g_l, b_l) and Beta (g_r, b_r).

Acknowledgments

We wish to thank Andrew Heathcote, Joshua Tenenbaum, and an anonymous reviewer for helpful comments, and Glen Smith, Michael Paradowski, Ryan Balzan, and Anthony Venning for assistance with the experiment. We would like to dedicate this article to Douglas Vickers, who passed away in October 2004. Douglas never saw this work, but we hope he would have enjoyed it.

References

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgments. *Perception & Psychophysics, 55,* 412–428.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology, 32,* 91–134.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic cognition approach to decision making. *Psychological Review, 100,* 432–459.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology, 41,* 260–274.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Gigerenzer, G., & Selten, R. (2001). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: MIT Press.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Jaynes, E. T. (2003). *Probability theory: The logic of science* (G. L. Bretthorst, Ed.). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (2000). *Choices, values, and frames.* New York: Cambridge University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.

Laming, D. R. J. (1968). *Information theory and choice-reaction time.* London: Academic.

Lee, M. D., & Corlett, E. Y. (2003). Sequential sampling models of human text classification. *Cognitive Science, 27,* 159–193.

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and "rational" models. *Psychonomic Bulletin & Review, 11,* 343–352.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77–105.

Myung, I. J., Balasubramanian, V., & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences, 97,* 11170–11175.

Myung, I. J., Forster, M., & Browne, M. W. (2000). A special issue on model selection. *Journal of Mathematical Psychology, 44,* 1–2.

Nickerson, R. S. (2004). *Cognition and chance: The psychology of probabilistic reasoning.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Peirce, C. S., & Jastrow, J. (1884). On small differences in sensation. *Memoirs of the National Academy of Sciences, 3,* 73–83.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109,* 472–491.

Price, P. C., & Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making, 17,* 39–57.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107,* 358–367.

Simon, H. A. (1976). From substantive to procedural rationality. In S. J. Latsis (Ed.), *Method and appraisal in economics* (pp. 129–148). London: Cambridge University Press.

Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology, 44,* 408–463.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology, 32,* 135–168.

Sniezek, J. A., & van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes, 84,* 288–307.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–640.

van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 582–600.

Vickers, D. (1979). *Decision processes in visual perception.* New York: Academic.

Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Non-Linear Dynamics, Psychology, and Life Sciences, 2,* 169–194.

Vickers, D., & Packer, J. S. (1982). Effects of alternating set for speed or accuracy on response time, accuracy, and confidence in a unidimensional discrimination task. *Acta Psychologica, 50,* 179–197.

Vickers, D., Smith, P. L., Burt, J., & Brown, M. (1985). Experimental paradigms emphasising state and process limitations: II. Effects on confidence. *Acta Psychologica, 59,* 163–193.

Wallsten, T. S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning, Memory & Cognition, 8,* 361–384.