Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48*(2), 87–106.

Birnbaum, M. H. (2004b). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes, 95,* 40–65.

Birnbaum, M. H. (2005a). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty, 31,* 263–287.

Birnbaum, M. H. (2005b). Three new tests of independence that differentiate models of risky decision making. *Management Science, 51,* 1346–1358.

Birnbaum, M. H. (2006). Evidence against prospect theories in gambles with positive, negative, and mixed consequences. *Journal of Economic Psychology, 27,* 737–761.

Birnbaum, M. H. (2007). *Testing heuristic models of decision making: Priority dominance, dimension integration, and dimension interaction.* Manuscript submitted for publication.

Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes, 71*(2), 161–194.

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semi-order. *Organizational Behavior and Human Decision Processes, 104,* 97–112.

Birnbaum, M. H., & LaCroix, A. R. (in press). Dimension integration: Testing models without trade-offs. *Organizational Behavior and Human Decision Processes.*

Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes, 67,* 91–110.

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17,* 49–78.

Birnbaum, M. H., & Sutton, S. E. (1992). Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processes, 52,* 183–215.

Birnbaum, M. H., & Veit, C. T. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception and Psychophysics, 15,* 7–15.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Choices without tradeoffs. *Psychological Review, 113,* 409–432.

Erev, I., Roth, A. E., Slonim, R. L., & Barron, G. (2002). *Combining a theoretical prediction with experimental evidence to yield a new prediction: An experimental design with a random sample of tasks.* Unpublished manuscript, Columbia University and Faculty of Industrial Engineering and Management, Techion, Haifa, Israel.

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

Lopes, L. L., & Oden, G. C. (1999). The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology, 43,* 286–313.

Mellers, B. A., Chang, S., Birnbaum, M. H., & Ordóñez, L. (1992). Preferences, prices, and ratings in risky decision making. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 347–361.

Rieger, M. O., & Wang, M. (in press). What is behind the priority heuristic: A mathematical analysis and comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review, XX,* xxx–xxx.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76,* 31–48.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5,* 297–323.

Received January 31, 2007
Revision received March 13, 2007
Accepted March 16, 2007 ■

Postscript: Rejoinder to Brandstätter et al. (2008)

Michael Birnbaum
California State University, Fullerton

Brandstätter, Gigerenzer, and Hertwig (2008, Figure 1) reanalyzed data of Erev, Roth, Slonim, and Barron (2002) to display a correlation between the accuracy of the priority heuristic and expected value (EV). From this correlation, they argued that people use two processes to choose between gambles: First, they act as if they compute ratios of EV and choose the gamble with the higher EV when this ratio exceeds a threshold, which the authors estimated to be two. Given this parameter, 56 of the 100 choices of Erev et al. were supposedly decided by EV. Second, people supposedly used the priority heuristic to make the remaining 44 choices. Whereas the priority heuristic alone had 15 errors predicting these 100 modal choices, the EV rule plus priority heuristic had six errors. But even with EV, a free parameter, and a lexicographic heuristic, this model does not fit as well as either the transfer of attention exchange (TAX) model or cumulative prospect theory (CPT), each of which had only one error when the same choices were reproduced. Because several models can fit these data almost equally well, I think it best to say that these data, like the Kahneman and Tversky (1979) data, are simply not diagnostic for comparing those models. Given the fact that they use an "as-if" (EV) model and estimate its parameter (two) post hoc, it seems odd that Brandstätter et al. continue to argue against "as-if" models and parameter estimation.

The EV plus priority heuristic model does not fit other, more diagnostic studies, as Brandstätter et al. (2008) acknowledged. To handle such data, they argued that each new failure of the EV plus priority heuristic should be taken as evidence for another heuristic. Among the additional heuristics they added to the theoretical stew are "dominance," "similarity," "toting up," "cancellation," "combination," and the "most-likely" heuristic. Consider how their "toting-up" heuristic was devised. Because EV plus priority heuristic is correct for fewer than half of the modal choices in Birnbaum and Navarrete (1998), Brandstätter et al. decided to replicate part of that study. There were 144 choices in the original study, including 112 in the main experimental designs that tested stochastic dominance and cumulative independence. For reasons I do not understand, Brandstätter et al. decided to examine only 54 of those 112 choices, so their new data did not allow tests of stochastic dominance or of upper or lower cumulative independence. Also puzzling were their decision to use a smaller number of participants than in the original study, their use of two new formats for the presentation of gambles, and other changes in procedure. Unlike the study by Birnbaum and Navarrete, every

choice in the replication study had a common probability-consequence branch, perhaps intended to help people to cancel such branches (as conjectured by Birnbaum & McIntosh, 1996). Given Brandstätter et al.'s new data, I found that the correlation between the group choice proportions of the new data and those of Birnbaum and McIntosh (1996) was 0.82 for the 54 choices they selected to replicate. The EV plus priority heuristic correctly predicted only 46% of the modal choices in their new study. A coin toss would have likely outperformed these heuristics for their new data, as it would have done in Birnbaum and Navarrete. Brandstätter et al. next focused their attention on 17 replicated trials from these 54 choices in which there was a common probability consequence branch and the other two probabilities were equal. For these cases, they proposed that people cancel the common branch and then apply a "toting up" heuristic, which correctly fit most of these 17 problems. According to this new heuristic, when probabilities are the same, people supposedly add up the consequences and choose the gamble with the higher total. However, this new heuristic does not resolve other choices in their replication (for which they entertained other heuristics), nor is it consistent with previously published data. For example, Birnbaum (1999, Table 3, Choice 2) reported a choice between ($100, 0.5; $0, 0.5) and ($45, 0.5; $35, 0.5). According to the toting-up hypothesis, the totals are $100 versus $80, and probabilities are equal, so a person should choose the gamble with the higher total. Instead, Birnbaum found that 69% of 124 college students and 60% of 1,224 Web recruits preferred the "safer" gamble with the smaller total. Eventually, 4,910 participants who received this same choice were tested using a variety of formats (Birnbaum, 2004b). Averaged over all participants, 65% preferred the gamble with the smaller toting value. Ironically, the priority heuristic would have been consistent with this modal choice had the toting-up heuristic not been added to the mix. The toting-up hypothesis also fails to fit other data with three-branch gambles. Birnbaum and McIntosh (1996) reported that 70% of 106 undergraduates chose a gamble with equal probabilities to win $2, $45, or $49, rather than a gamble with equal chances to win $2, 11, or $97, even though the total of the former is $96, whereas the latter adds up to $110. Many other examples that significantly violate the EV plus toting-up plus priority heuristics can be found in Birnbaum (2004a, 2004b, 2007), Birnbaum and Chavez (1997), Birnbaum and McIntosh, and Birnbaum and Navarrete. In sum, although the new heuristic predicts the majority of the 17 choices it was designed to predict, the EV plus toting-up plus priority heuristic does not predict other, previously published data.

Similarly, a "similarity heuristic" was modified by Brandstätter et al. (2008) to predict violations of stochastic dominance reported by Birnbaum and Navarrete (1998), which violate the EV plus priority heuristic. This heuristic contradicts the priority heuristic because it assumes that people do look at consequences on the middle branches of three-branch gambles and use large differences on that branch when all other differences are below the thresholds in the priority heuristic. Although this heuristic fits the violation of stochastic dominance that it was devised to fit, it does not work in fitting other violations of stochastic dominance (see Birnbaum, 2005). In addition, it does not explain why violations of stochastic dominance are found in judgment, as well as in choice studies. Surprisingly, Brandstätter et al. also endorsed the editing principles of Kahneman and Tversky (1979), including combination and cancellation, despite a large body of evidence contradicting those two principles (see the review in Birnbaum, 2006).

Brandstätter et al. (2008) suggested that their model-fitting contests should be called tests of "predictions." I find it odd to say that one has "predicted" previously published data, because the term "prediction" means to *say in advance*. There are two ways to compute the fit of models to previous data: One can estimate best-fit parameters for both theories under investigation, or one can use parameters chosen in advance. These different types of analyses were described and carried out in Birnbaum and Navarrete (1998), for example, on both individual and aggregate data. These different methods can lead to different conclusions, so one should examine them all to be on secure footing. The method that Brandstätter et al. argued should be the *only* method of analysis is one that has been called the "tool of the Devil" when used to diagnose models because it can easily lead to wrong conclusions (Birnbaum, 1973, 1974). Brandstätter et al. did not address the hazards of comparing fit of models using prior parameters; they simply argued that the term "prediction" should be used for this method and "fitting" should be used for more conventional methods. Whether one uses a correlation coefficient, sum of squares, likelihood, or percentage correct, comparisons of fit (even if called "predictions") with a priori parameter values easily lead to wrong conclusions. The priority heuristic does not outperform its rivals when assessed by more appropriate and traditional methods of analysis (e.g., as in Figures 1 and 2 of Birnbaum, 2008), nor does this heuristic do well when applied to data from experiments that are more diagnostic than those reviewed by Brandstätter et al.

Brandstätter et al. (2008) argued that models with free parameters are difficult to test. This statement applies only if we restrict our attention to post hoc fitting ("prediction") contests with nondiagnostic data. The statement is false, however, in an experimental field such as psychology, in which it is possible to devise and test critical properties of models. For example, despite its many free parameters, CPT implies that people should always conform to first-order stochastic dominance, except for random error. There is no set of parameters that allows CPT to imply that A = ($96, 0.55; $90, 0.35; $12, 0.1) should be preferred to B = ($96, 0.9; $14, 0.05; $12, 0.05). Like CPT, the latest combination of heuristics also predicts that the majority should prefer B to A because B dominates A, B has a much higher probability of the highest consequence, and B has the higher EV. The latest six-heuristic model ("no-conflict solution," plus as-if EV, plus "similarity," plus "cancellation," plus "toting up," plus priority heuristic) described by Brandstätter et al. fails to "predict" that people choose B over A. Nevertheless, 61% of 428 subjects in Birnbaum (2005) showed this preference. Similarly, both CPT and TAX must satisfy transitivity of preference, which, if violated, would have been an argument for the priority heuristic over these parametric models. Birnbaum and Gutierrez (2007) tested transitivity where predicted by the priority heuristic and found that the vast majority of participants were consistent with transitivity. The priority heuristic was correct in predicting only 30% of the modal choices in their studies.

Brandstätter et al. (2008) appear to have adopted the position that if they can find a heuristic that can "predict" (fit) previously published data, that model is true for those cases in which its "predictions" are correct. When that model fails to predict other data, they infer that there must be additional processes involved.

This position seems rather complicated because it involves at least one preliminary decision mechanism that uses "triggering conditions" to determine which decision heuristics to use. Without a doubt, a new heuristic can be added or an old one revised to handle violations of their latest combination of heuristics. Although the general position they have adopted appears to be one that cannot be refuted in principle, the particular combination of heuristics that they have so far advocated does not appear to be accurate beyond the sets of data in which those heuristics were devised.

References

Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin, 79,* 239–242.

Birnbaum, M. H. (1974). Reply to the Devil's advocates: Don't confound model testing and measurement. *Psychological Bulletin, 81,* 854–859.

Birnbaum, M. H. (1999). Testing critical properties of decision making on the Internet. *Psychological Science, 10,* 399–407.

Birnbaum, M. H. (2004a). Causes of Allais common consequence paradoxes: An experimental dissection. *Journal of Mathematical Psychology, 48*(2), 87–106.

Birnbaum, M. H. (2004b). Tests of rank-dependent utility and cumulative prospect theory in gambles represented by natural frequencies: Effects of format, event framing, and branch splitting. *Organizational Behavior and Human Decision Processes, 95,* 40–65.

Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty, 31,* 263–287.

Birnbaum, M. H. (2006). Evidence against prospect theories in gambles with positive, negative, and mixed consequences. *Journal of Economic Psychology, 27,* 737–761.

Birnbaum, M. H. (2007). *Testing heuristic models of decision making: Priority dominance, dimension integration, and dimension interaction.* Manuscript submitted for publication.

Birnbaum, M. H. (2008). Evaluation of the priority heuristic as a descriptive model of risky decision making: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review, 115,* 253–262.

Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes, 71*(2), 161–194.

Birnbaum, M. H., & Gutierrez, R. J. (2007). Testing for intransitivity of preferences predicted by a lexicographic semiorder. *Organizational Behavior and Human Decision Processes, 104,* 97–112.

Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes, 67,* 91–110.

Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty, 17,* 49–78.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2008). Risky choice with heuristics: Reply to Birnbaum (2008); Johnson, Schulte-Mecklenbeck, & Willemsen (2008); and Rieger & Wang (2008). *Psychological Review.*

Erev, I., Roth, A. E., Slonim, R. L., & Barron, G. (2002). *Combining a theoretical prediction with experimental evidence to yield a new prediction: An experimental design with a random sample of tasks.* Unpublished manuscript, Columbia University and Faculty of Industrial Engineering and Management, Techion, Haifa, Israel.

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