In press: Hills, T., & Hertwig, R. (In press). Two distinct exploratory behaviors in decisions from experience: Comment on Gonzalez & Dutt, 2011

Two Distinct Exploratory Behaviors in Decisions From Experience: Comment on Gonzalez & Dutt, 2011

Thomas T. Hills

Department of Psychology, University of Warwick

&

Ralph Hertwig

Cognitive and Decision Sciences, University of Basel

Word count (main text): 2226

Address correspondence to:
Thomas Hills
University of Warwick
Department of Psychology
Gibett Hill Road
Coventry CV4 7AL, UK
Phone: +44-(0) 24-7657-5527
E-mail: t.t.hills@warwick.ac.uk

Abstract

Gonzalez and Dutt (2011) recently reported that trends during sampling, prior to a consequential risky decision, reveal a gradual movement from exploration to exploitation. That is, even when search imposes no immediate costs, people adopt the same pattern manifest in costly search: early exploration followed by later exploitation. From this isomorphism the authors conclude that the same cognitive mechanisms underlie the control of sampling in two experimental paradigms employed to investigate decisions from experience—the sampling paradigm implementing costless search and the repeated choice paradigm implementing costly search. We show that this is a misinterpretation of the data resulting from drawing inferences about cognitive processes from data aggregated across individuals. Because of an inverse relationship between sample size and the propensity to explore, aggregating across individuals produces a pattern where exploration is gradually replaced by exploitation. On an individual level, however, there is no general reduction in exploration during the sampling in the sampling paradigm. We list ensuing problems for the instance-based learning model Gonzalez and Dutt present to explain the similarities between sampling and repeated decisions from experience.

Keywords: Decisions from experience, risky choice, exploration, exploitation, search, data aggregation

Risky decisions based on experience often come in two variants. Either experience comes before the consequential choice based on exploratory—yet non-consequential—sampling of the environment or, alternatively, experience comes with the choice, as the result of making consequential decisions and thus learning from one's successes and failures. In the first variant, *exploration* (e.g., obtaining information) and *exploitation* (obtaining reward) are two separate processes, and the agent's only initial objective is to explore the environment in order to find out which of her actions is most instrumental in obtaining future rewards. Consequential exploitation follows costless exploration at a time of the agent's choosing, and thus does not require the agent to find a balance between the opportunity costs of both objectives. In the second variant, the sampled outcomes simultaneously provide reward *and* information to the agent, and thus she faces the exploration–exploitation dilemma: "The agent has to *exploit* what it already knows in order to obtain reward, but it also has to *explore* in order to make better action selections in the future" (Sutton & Barto, 1998, p. 4). The agent thus has to find a trade-off between both.

In the laboratory, these two types of risky decisions from experience have been abstracted into what Hertwig and Erev (2009) have called, respectively, the *sampling paradigm* and the *partial-feedback paradigm* (Gonzalez and Dutt's, 2011, refer to the latter as the *repeated-choice paradigm*; henceforth, we adopt their language in order to avoid confusion). The latter imposes the exploration–exploitation dilemma on the agent; the former does not (see Gonzalez & Dutt, 2011, p. 525, for a detailed description of the paradigms). Because of this difference researchers have commonly proposed disparate cognitive models to account for the choices obtained in each of the paradigms (see Erev, Ert, Roth, et al., 2010, and Gonzalez & Dutt, 2011, pp. 528-530). Challenging this theoretical divide, Gonzalez and Dutt (2011) recently proposed an important theoretical framework, the *instance-based learning* model (IBL model), with the goal of explaining choice and search in both paradigms

using the same cognitive mechanisms (for a detailed description see Gonzalez and Dutt, 2011, p. 526-527).

The IBL model builds in some important respects on the ACT-R cognitive architecture (Anderson & Lebiere, 1998, 2003). Specifically, it assumes that a choice (given that the previous choice is not automatically repeated) represents the selection of the option with the highest utility (blended value). An option's blended value is a function of its associated outcomes and the probability of retrieving corresponding instances from memory. Memory retrieval depends on memory activation, which, in turn, is a function of the recency and frequency of the experience. The IBL model is particularly attractive because in the sampling paradigm it, unlike many other models that have been proposed, "predicts not only the final consequential choice but also the sequence of sampling selection" (Gonzalez & Dutt, 2011, p. 529), and because it offers a single learning mechanism (leading up to an instance's activation) that underlies both sequential choice and process behavior in the sampling and the repeated-choice paradigms.

The goal of the present comment is to show that Gonzalez and Dutt's (2011) commendable attempt to explain both paradigms in terms of the same mechanisms faces several potentially serious problems. We hope that their framework may ultimately be viable but in its current form the IBL model appears to invite inaccurate inferences about individual search behavior.

Is the Exploration–Exploitation Dynamic Indeed the Same in Both Paradigms?

The key issue concerns the relationship between exploration and exploitation in both paradigms. Based on their analyses, Gonzalez and Dutt (2011) concluded that what seemingly separates the paradigms may in fact unite them, the temporal dynamics of exploration and exploitation. Specifically, they observed that "in both paradigms, the A-rate decreases over an increased number of samples or trials. Furthermore, the same IBL model calibrated in one paradigm with the same parameters predicts the A-rate of the other

paradigm" (pp. 538-539). The A-rate or alternation rate is a measure of the amount of exploration, and denotes the proportion of times that an individual moves from choosing one option to choosing the other option during periods of sampling (in the sampling paradigm; see also Hills & Hertwig, 2010) and repeated-choices (in the repeated-choice paradigm). This measure of exploration is commonly used and is typically found to decay over time in repeated-choice tasks as individuals move from exploration to exploitation (e.g., Yechiam, Busemeyer, Stout, & Bechara, 2005). According to Gonzalez and Dutt's (2011) empirical and IBL model analyses (see their Figures 2 and 3), respondents in the sampling paradigm behave like respondents in the repeated-choice paradigm: Over time, exploitation supersedes exploration, and thus people at this non-consequential search stage eventually act as if they adopted an exploitative goal.

This isomorphism in non-consequential search and consequential choice would indeed draw both paradigms nearer. However, the move from exploration to quasi-exploitation in the search sequence for the sampling paradigm mischaracterizes what many individuals do. In a nutshell the problem is the following: First, individuals vary in their search length in the sampling paradigm. Second, as a function of search length the A-rate varies considerably. Third, when A-rate is analyzed as a function of a normalized search length the sampling paradigm reveals, on average, a strikingly *constant* level of exploration over the entire duration of the exploratory sampling phase.

Individual Versus Average Explorative Behavior in the Sampling Paradigm

Gonzalez and Dutt (2011) analyzed two data sets obtained in the sampling paradigm: the data collected by Hertwig et al. (2004) and the data collected in the sampling condition of the *Technion Prediction Tournament* (TPT; Erev, Ert, Roth, et al., 2010). Figure 1 replots the average decreasing A-rate that they found in the data sets. Inferring from this pattern that individuals commonly move from exploration to exploitation, however, is wrong. The reason lies in the inverse relationship between search length and A-rate. As shown in the panels of

Figure 2 (left and middle), individuals' total sample sizes A-rate are significantly long-tailed (Shapiro-Wilks tests are $p < 0.001$ for all variables). Therefore, we employed a log transformation to compute the correlations (Shapiro-Wilks tests after the log transformation are $p > 0.1$ for both variables in both data sets). The correlation coefficients between the log of total sample size and log A-rate reveals a significant negative correlation for both data sets (Hertwig et al., 2004, data: Pearson correlation = -.38, $t(48) = -2.82$, $p < 0.01$; TPT data: Pearson correlation = -0.54, $t(78) = -5.72$, $p < 0.001$). The negative correlations across individuals for both data sets are depicted in Figure 2 (right panels).

[Figures 1 and 2]

Given this negative relationship, aggregating the A-rate across individuals for different absolute numbers of samples (as done in Figure 1) means that those individuals whose total sample size is relatively small (and whose A-rate is relatively high) drop out of the analysis, leaving only those people behind who sample more and alternate less. Consequently, the (possibly) erroneous impression arises that exploration is superseded by exploration on an individual level. One way to deal with the inverse relationship of total sample size and A-rate is to normalize search sequences, and then analyze the aggregate A-rate. Figure 3A plots the resulting A-rate in both data sets as a proportion of total samples. Now, the conclusion is that, on average, the A-rate is quite *constant* across the sampling sequences, and thus qualitatively different from the declining exploration rate obtained in the repeated-choice paradigm (see Gonzalez & Dutt's, 2011, Figure 2A and 3A). Note that calculating an aggregated A-rate in the repeated-choice paradigm is less problematic because the number of trials is held constant across individuals (see, for instance, Erev, Ert, Roth, et al., 2010). We found the same constant average A-rate (using normalized search sequences) in several other data sets from the sampling paradigm (Hau, Pleskac, Kiefer, & Hertwig, 2008; Hertwig & Pleskac, 2010; Ungemach, Chater, & Stewart, 2009).

[Figure 3]

Figure 3A, however, still plots an aggregated A-rate. To reveal individual trends in exploration we calculated the A-rate for the first 25% and last 25% of their sampling sequence for each person. Figure 3B plots the initial and final A-rates. If indeed exploration eventually superseded exploitation in the sampling paradigm all data points should cluster in the triangle below the diagonal. That is not the case. Instead, a minority of 10% (Hertwig et al., 2004) and 5% of people (Erev, Ert, Roth, et al., 2010) has constant initial and final A-rates. The remaining participants fall in about equal size classes: In the Hertwig et al. (2004) data set, 23 show a reduction in A-rate, whereas 22 show an increase in A-rate. In the TPT data set, 18 show a reduction, and 20 show an increase.

Ensuing Problems for the IBL Model

These findings highlight the dangers of drawing inferences from the average A-rate to individual search behavior, and they matter for the IBL model in its current form. If one accepts the constant average A-rate calculated across the normalized sequence as a more veridical representation of the aggregate and a better approximation of the individual behavior, relative to the declining A-rate, then at least the following problems arise: First, the inferred isomorphism in the A-rate disappears, and thus makes both paradigms much less similar than suggested by Gonzalez and Dutt's (2011) analysis. Second, the probability of retrieving instance *i* from memory, and its activation (see Equations 3 and 4 in Gonzalez & Dutt, 2011) cannot change as predicted for both the repeated-choice and the sampling paradigms. Specifically, in the sampling paradigm, there would appear to be no effect of the magnitude of the blended value for each option (*V* in Equation 2 from Gonzalez & Dutt, 2011), because the probability of choosing a specific option does not change systematically over the sampling interval. If the blended value of one option grew larger than the other option—as it must if the final choice is to be predicted accurately—then would lead to less exploration over time as the sampling of one choice is favored over the other. However, as

noted above, the data from the sampling paradigm does not show a systematic reduction in exploration at the individual level. This is a drastic difference from the choice behavior observed in the repeated-choice paradigm.

Third, the IBL model involves the adjustable parameter *pInertia*, which determines whether the choice made in the previous trial is repeated or not. If *pInertia* equals 1, then the IBL model will always repeat the last choice, or in other words, it will predict no exploration. Gonzalez and Dutt (2011) concluded that "in both paradigms, the A-rate decreases over an increased number of samples or trials", and that "results suggest that in both paradigms, humans move gradually from the exploration of options to their exploitation using the same cognitive mechanisms for the sequential selection of alternatives" (p. 539). The *pInertia*, estimated from human data, however, does not support this conclusion: They widely diverge across paradigms (p. 539), inconsistent with the conclusion that behavior in both paradigms is "equivalent … at the sequential process (A-rate) level".[1]

The fourth problem the IBL model faces is that in the sampling paradigm it fails to predict the inverse correlations between A-rate and total sample size (see Figure 2, right-most panels). In order to account for total sample size, the IBL model randomly draws a value from distributions fitted to the empirically observed sample sizes in the Hertwig et al. (2004) and the TPT data set (Erev, Ert, Roth, et al., 2010) (Gonzalez & Dutt, 2011, p. 527). Consequently, the IBLT modeling uses sample size distributions that have no relationship to the equations that generate alternations. Ergo, according to the IBL model there is no relationship between sample size and alternations.

Conclusion

The IBL model is a very commendable attempt to explain the behavior in two experimental paradigms using the same cognitive mechanisms. And, indeed, the sampling

[1] Their divergence, however, is unsystematic across two calibration sets. In one set, the calibration resulted in *pInertia* equals 0.22 and 0.48 for the sampling and repeated-choice paradigm, respectively. In the other calibration set, however, the order is reversed and *pIntertia* equals 0.63 and 0.09 in the sampling and repeated-choice paradigms, respectively. We do not know why these values vary so widely—but to the extent that *pInertia* is correlated with A-rate, their divergence does not support the conclusion that behavior in both paradigms is "equivalent".

and the repeated-choice paradigm have been found to produce surprisingly similar choice patterns, leading to a similar kind of description–experience gap (see Hertwig & Erev, 2009). This similarity, however, is not replicated on other cognitive dimensions. In contrast to Gonzalez and Dutt's (2011, p. 539) conclusion, there appears to be no general move from exploration of options to their exploitation in the sampling paradigm. Their conclusion results from making inferences about individual behavior using data aggregated over individuals. Dealing with similar issues, Estes and Maddox (2005) highlighted the "danger … that individual differences among subjects with respect to values of a model's parameters may cause averaging to produce distorted inferences about true patterns of individual performance and the cognitive processes underlying them" (p. 403).

But let us not throw the baby out with the bathwater. The IBL model can predict final decisions in the sampling paradigm as well as or better than any other proposed model. It can also predict choices in the repeated-choice paradigm better than other models. This is obviously an excellent model on the level of choice, and some of its building blocks (e.g., the activation mechanism) have been demonstrated to be instrumental in successfully modeling a wide range of behaviors (e.g., Anderson & Lebiere, 1998, 2003; Gonzalez et al., 2010; Gonzalez & Lebiere, 2005). However, the model in its present form does not get the relative balance of exploration and exploitation in non-consequential search right.

References

Anderson, J. R. & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Anderson, J. R. & Lebiere, C. L. (2003). The Newell test for a theory of cognition. *Behavioral & Brain Science, 26*, 587-637.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., … Lebiere, C. (2010). A choice prediction competition: Choices from experience and description. *Journal of Behavioral Decision Making, 23,* 15-47.

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review, 12,* 403-408.

Gonzalez, C., Best, B. J., Healy, A. F., Bourne, L. E. Jr, & Kole, J. A. (2010). A cognitive modeling account of simultaneous learning and fatigue effects. *Journal of Cognitive Systems Research, 12,* 19–32.

Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating sampling and repeated decisions from experience. *Psychological Review, 118,* 523-551. doi: 10.1037/a0024558.

Gonzalez, C., & Lebiere, C. (2005). Instance-based cognitive models of decision making. In D. Zizzo & A. Courakis (Eds.), *Transfer of knowledge in economic decision-making* (pp. 148-165). New York, NY: Palgrave Macmillan.

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description–experience gap in risky choice: the role of sample size and experienced probabilities. *Journal of Behavioral Decision Making, 21,* 493-518.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15,* 534-539.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences*, *13*, 517-523.

Hertwig, R., & Pleskac, T. J. (2010). Decisions from experience: Why small samples? *Cognition*, *115*, 225-237.

Hills, T. T., & Hertwig, R. (2010). Information search in decisions from experience: Do our patterns of sampling foreshadow our decisions? *Psychological Science, 21,* 1787–1792. doi:10.1177/0956797610387443.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science, 20,* 473–479.

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science*, *16,* 973-978.

Acknowledgements

We thank Laura Wiles for editing the manuscript. This research was supported by grants from the Swiss National Science Foundation to the first (100014 130397/1) and second author (100014 126558).

Figure captions

Figure 1

The A-rate (alternation rate between options) for the two sampling paradigm data sets analyzed by Gonzalez and Dutt (2011): Hertwig et al. (2004) and the Technion Prediction Tournament data set (TPT; Erev, Ert, Roth, et al., 2010). The A-rate is aggregated first within individuals and then over individuals at each sample size.

Figure 2

Histograms and correlations for total number of samples taken and the A-rate for the two sampling paradigm data sets analyzed by Gonzalez and Dutt (2011): Hertwig et al. (2004) and the Technion Prediction Tournament data set (TPT; Erev, Ert, Roth, et al., 2010). Lines in figures on the right-hand-side represent the best fitting regression lines between log of total number of samples and the log of the overall A-rate.

Figure 3

The normalized A-rate (alternation rate between options) for the two sampling paradigm data sets analyzed by Gonzalez and Dutt (2011): Hertwig et al. (2004) and the Technion Prediction Tournament data set (TPT; Erev, Ert, Roth, et al., 2010). (A) The A-rate normalized by proportion of the sampling interval in 10% bins. (B) The A-rate for the first 25% and the last 25% of the normalized sampling sequences. Circles on the diagonals represent people whose initial and final A-rates are identical.

Figure 1

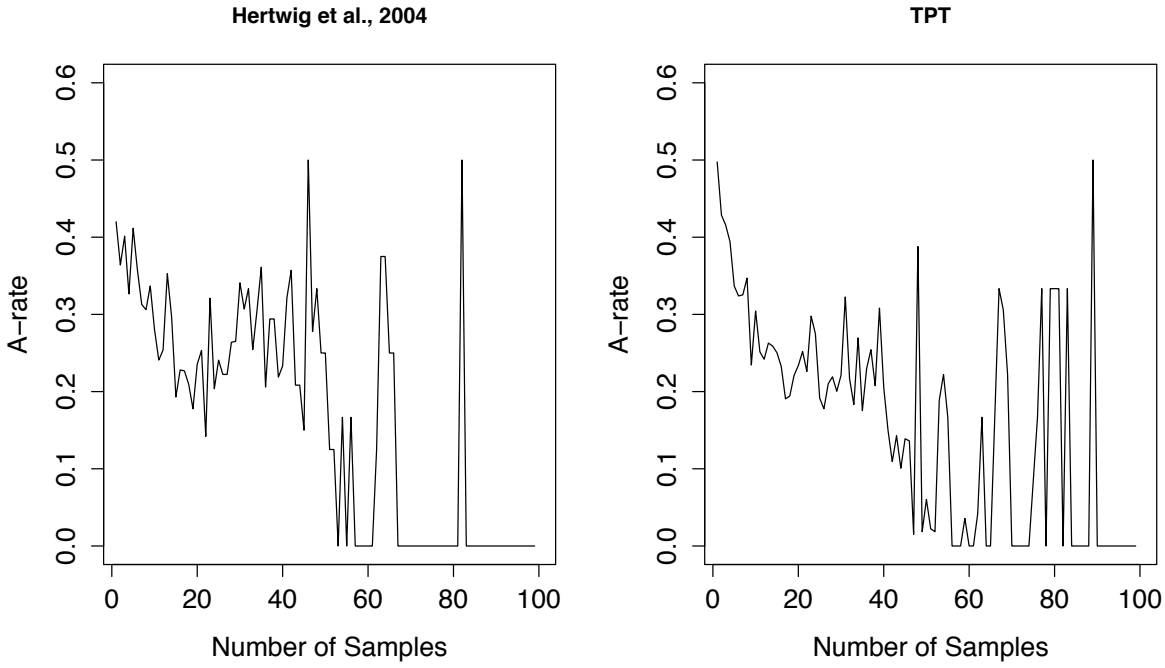

Hertwig et al., 2004
TPT
A−rate
Number of Samples
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
20
40
60
80
100

Figure 2

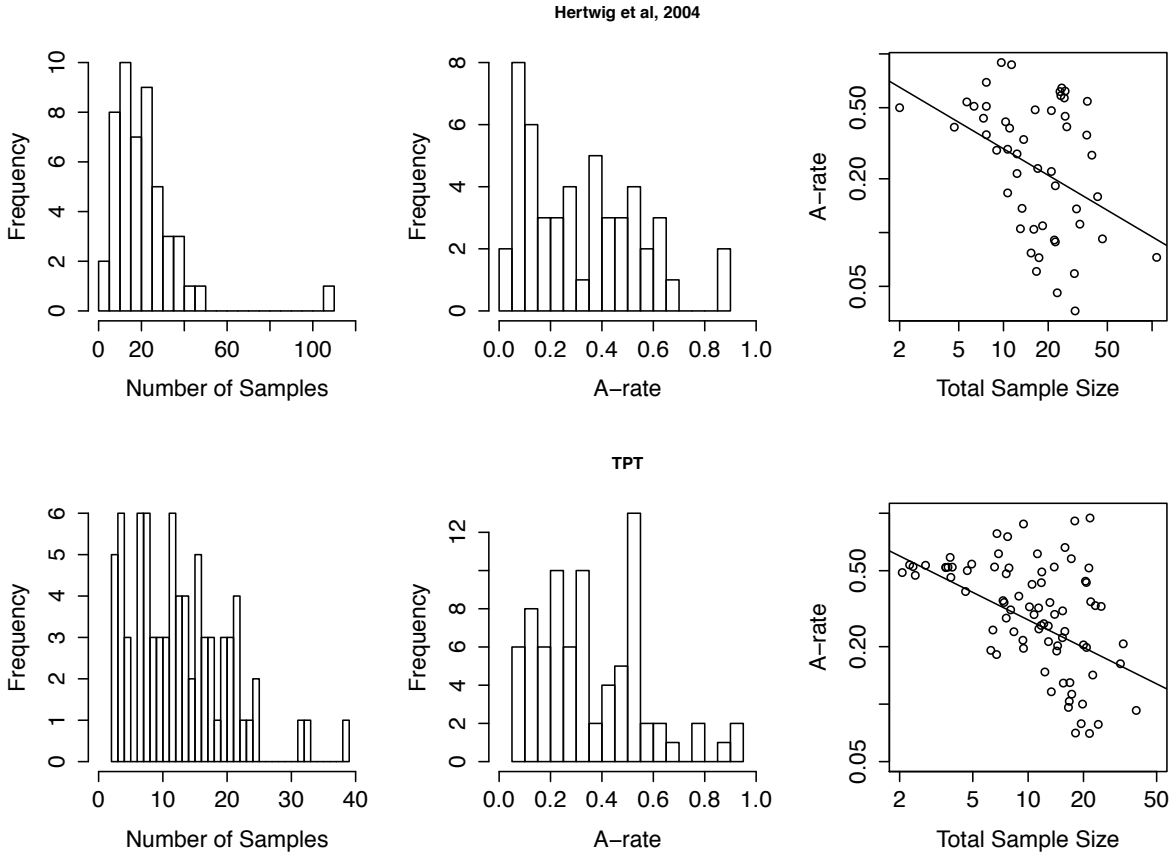
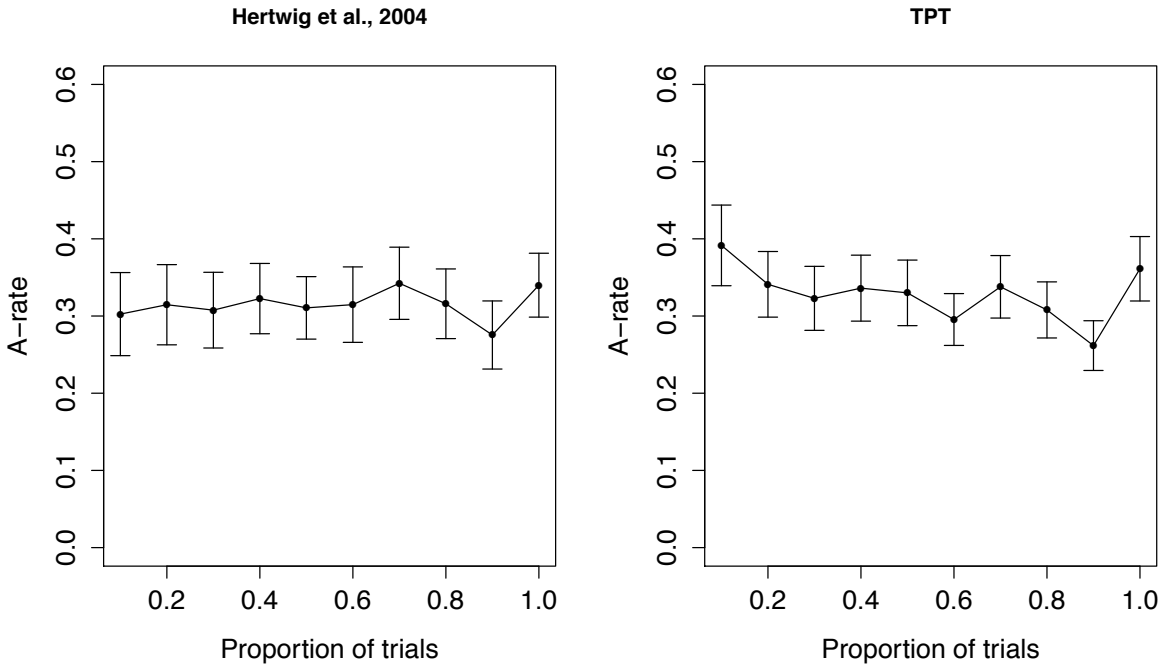
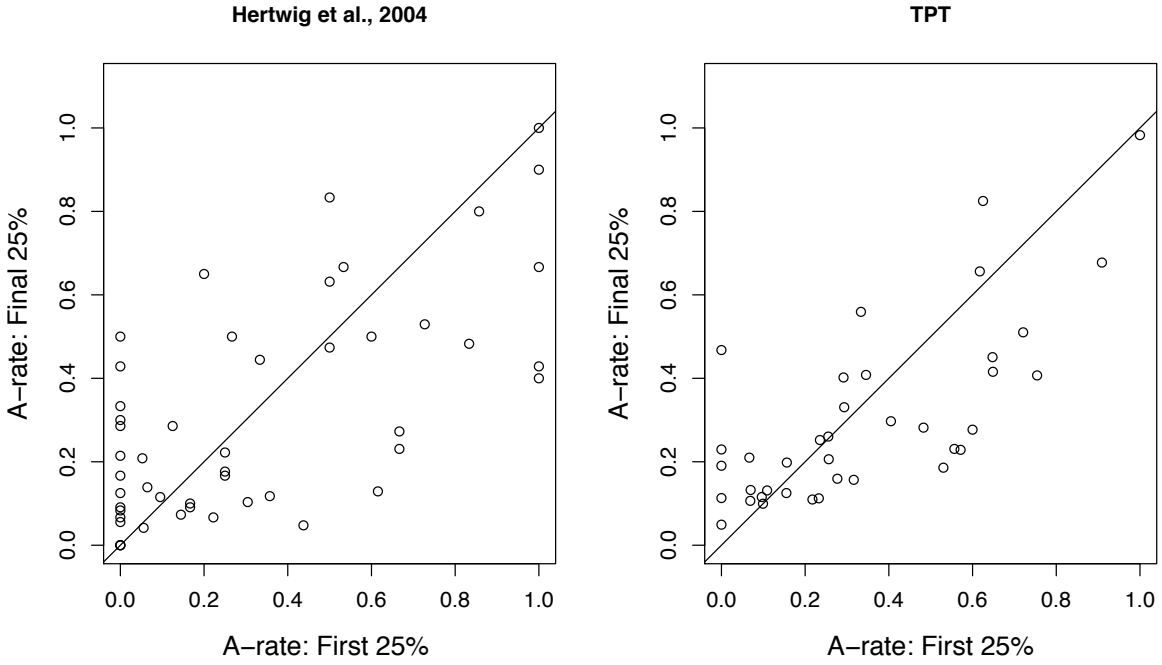

Hertwig et al, 2004
Frequency
Number of Samples
A−rate
Total Sample Size
TPT
Frequency
Number of Samples
A−rate
Total Sample Size

Figure 3

A


Hertwig et al., 2004
TPT
A−rate
Proportion of trials
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.2
0.4
0.6
0.8
1.0

B


Hertwig et al., 2004
TPT
A−rate: Final 25%
A−rate: First 25%
0.0
0.2
0.4
0.6
0.8
1.0