

Rational Learning and Information Sampling: On the “Naivety” Assumption in Sampling Explanations of Judgment Biases

Gaël Le Mens

Universitat Pompeu Fabra and Barcelona Graduate School of
Economics

Jerker Denrell

University of Oxford

Recent research has argued that several well-known judgment biases may be due to biases in the available information sample rather than to biased information processing. Most of these sample-based explanations assume that decision makers are “naive”: They are not aware of the biases in the available information sample and do not correct for them. Here, we show that this “naivety” assumption is not necessary. Systematically biased judgments can emerge even when decision makers process available information perfectly and are also aware of how the information sample has been generated. Specifically, we develop a rational analysis of Denrell’s (2005) experience sampling model, and we prove that when information search is interested rather than disinterested, even rational information sampling and processing can give rise to systematic patterns of errors in judgments. Our results illustrate that a tendency to favor alternatives for which outcome information is more accessible can be consistent with rational behavior. The model offers a rational explanation for behaviors that had previously been attributed to cognitive and motivational biases, such as the in-group bias or the tendency to prefer popular alternatives.

Keywords: learning, sampling, judgment biases, rational analysis

Supplemental materials: <http://dx.doi.org/10.1037/a0023010.supp>

As emphasized by Herbert Simon more than 50 years ago (Simon, 1955, 1956), understanding decision making requires examining both the heuristics used by organisms and the structure of the information provided by their environment. Much subsequent research in psychology has focused on the properties of heuristics (Tversky & Kahneman, 1974) used by individuals to form judg-

ments. In this line of research, explanations of systematic biases in judgment and decision making are typically sought out in flawed information processing. The emphasis is on how decision makers with access to unbiased and representative samples could end up with biased beliefs and make biased judgments.

The view that biased judgments are uniquely driven by flawed processing of information by the mind of individuals has recently been challenged by a number of scholars. They have focused on the second aspect of Simon’s lemma: the role of the environment, and how it can sometimes produce biased samples of information, unbeknown to the decision maker (see Fiedler & Juslin, 2006, for a review). This sampling perspective has produced alternative explanations to several prominent biases in judgment, including illusory correlations (Denrell & Le Mens, 2008; Fiedler, 2000), overconfidence (Juslin, Winman, & Hansson, 2007), in-group bias and devaluation of minorities (Denrell, 2005; Fiedler, 2000), social influence (Denrell & Le Mens, 2007), and risk aversion (Denrell, 2007; Hertwig, Barron, Weber, & Erev, 2004; March, 1996).

In contrast to the heuristics and biases program, sampling explanations assume that decision makers can accurately assess the properties of the samples of information available to them. The focus is instead on how biased judgment can emerge when decision makers extrapolate from the typically small and biased sample of information available to them. According to Fiedler and Juslin (2006), decision makers behave as if they were naive intuitive statisticians. That is, people behave as if they were statisticians, accurately summarizing the data but being naive about the potential biases in the data available to them.

Gaël Le Mens, Department of Economics and Business, Universitat Pompeu Fabra, Barcelona, Spain, and Barcelona Graduate School of Economics, Barcelona, Spain; Jerker Denrell, Saïd Business School, University of Oxford, Oxford, England.

Both authors contributed equally to this work. This article benefited from comments by participants at presentations at the Barcelona Economic Decision Group (October 2009), INSEAD (January 2009), Hong Kong University of Science and Technology (November 2008), London Business School (January 2009), Massachusetts Institute of Technology (February 2009), Universitat Pompeu Fabra (December 2008), University of Oxford (December 2008), University of Warwick (November 2009), Yale University (December 2008), and the Society for Judgment and Decision Making Conference (November 2008). Gaël Le Mens was supported by a Stanford Graduate School of Business Fellowship and by a Juan de la Cierva Fellowship from the Spanish Ministry of Science and Education. We are grateful to Jon Bendor, Glenn Carroll, Ido Erev, Mike Hannan, Robin Hogarth, Matt Jackson, Sunil Kumar, Laurent Le Brusquet, Kristóf Madarász, John Roberts, Tomasz Sadzik, Karl Schlag, Andy Skrzypacz, Damien Stehle, Pete Veinott, and George Wu for insightful discussions and comments.

Correspondence concerning this article should be addressed to Gaël Le Mens, Universitat Pompeu Fabra, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain. E-mail: gael.le-mens@upf.edu

The sampling approach has illustrated how basic judgment biases can be parsimoniously explained by assuming a common source of bias. Nevertheless, because sampling explanations assume that decision makers are naive with respect to the biases in the data available to them, sampling explanations continue to attribute judgment biases to biased or less than rational information processing, albeit the information processing bias is now located elsewhere.

In this article, we demonstrate that the naivety assumption is not necessary for the validity of sampling explanations. This result is important because it suggests that sampling explanations can be even more parsimonious than previously acknowledged. Furthermore, it potentially increases the scope and applicability of sampling explanations to settings where they were commonly thought inapplicable. To develop our argument, we adopt the rational analysis approach, initiated by Anderson (1990) and used in several influential contributions (e.g., Oaksford & Chater, 1994, 1998, 2007; Tenenbaum & Griffiths, 2001). We assume that decision makers (1) process the information in the available sample perfectly, according to Bayes's theorem, (2) possess complete knowledge of how the available sample has been generated, and (3) do not have any cognitive limitation as to the corrections they can apply to the available sample before drawing inferences. We show that even under these assumptions, judgments can be systematically biased.

We propose that this can occur when the goal of the decision maker is distinct from that of developing accurate knowledge. Specifically, we focus on situations where information search is *interested* rather than *disinterested* (Oaksford & Chater, 1994, 2007). We analyze an extension of the experience sampling model proposed by Denrell (2005) and show that even if decision makers are aware of the sample bias assumed in this model, the result will nevertheless be systematically biased judgments.

We consider a setting where the decision maker has to make a sequence of choices between two actions. Her goal is to maximize the expected total payoff she obtains over the complete sequence of choices. She observes the payoff generated by the first action in every period, no matter what action she chooses. However, she only observes the payoff generated by the second action if she selects it. In other words, the decision maker obtains incidental outcome information for the first action even if she does not select this action. However, for the second action, access to outcome information is selective: She can only observe the outcome of this action if she selects it. Here, we show that even a rational decision maker who attempts to solve this task will develop systematically biased beliefs. More precisely, we demonstrate that rational choice and information processing lead to a tendency to prefer the action with systematic access to incidental information to the action with selective feedback, even if both actions are equally likely to be the best.

This asymmetry is surprising because observations of both actions are independent realizations of the underlying payoff distributions. Besides, the decision maker possesses correct prior beliefs and is assumed to process information rationally. Furthermore, she not only knows the structure of the environment (i.e., how the information about the two actions is produced and accessed) but even controls her sampling behavior.

To develop an intuition of how the asymmetry in estimates can emerge, consider the following example. Imagine you form an

impression about a junior scholar who just joined your department and suppose your goal is to have interesting research conversations with your colleagues. Even if you did not find your first interaction with this researcher interesting and you do not interact much with him or her subsequently, you will almost inevitably learn more about his or her achievements through discussions with your colleagues. In that situation, the sample of information you obtain about your junior colleague does not uniquely depend on your goal, because you also obtain incidental information about him or her from other members of the department. Access to such incidental information implies that mistaken negative impressions may be corrected even if you avoid your junior colleague.

Compare this impression formation process with how you form an impression of a junior scholar at a conference. Suppose this is the first time you attend a talk by this researcher and that you do not know much about his or her research. The presentation is poorly organized, the theory seems vague, and the methodology is weak. You are likely to leave the talk with a negative impression of the presenter. Unless you are especially interested in the topic, you will probably not read the article on which the talk was based or attend other talks by this researcher. As a result of this avoidance behavior, your negative impression about the competence of the researcher is likely to persist, even if this researcher is in fact skilled. Overall, this implies that you will have a tendency to develop a more positive impression of the junior scholar in your department than of the junior scholar you met at the conference.

As demonstrated by Denrell (2005), it is possible to derive this conclusion formally if belief formation and choice are assumed to follow well-known heuristic models, such as the delta rule and the Luce choice rule (Luce, 1959). Because Denrell's model assumed bounded rationality, however, it is not clear that the same asymmetry would emerge if decision makers were rational. Intuitively, one might believe that rational decision makers should be able to correct for any bias caused by their own sampling behavior. After all, a rational decision maker who controls her own sampling behavior should understand the possible biases caused by her sampling choices. The main contribution of this article is to show that even rational decision makers, who know the structure of the environment and update their beliefs using Bayes's rule, would exhibit the same tendency to favor the first alternative, for which information is more accessible.

In addition to proving that the naivety assumption is not necessary for the validity of sampling explanations, this result has important empirical implications. In particular, it illustrates that a tendency to favor alternatives for which information is more accessible can be consistent with rational behavior. The model thus offers a rational explanation for behavior that had previously been attributed to flawed information processing and reasoning, such as the tendency for people to have more positive impressions of those they are frequently exposed to.

The rest of the article is organized as follows. In the first section, we analyze a simple learning model, and we demonstrate the emergence of a preference for the alternative with access to incidental information using computer simulations. In the second section, we prove that this pattern holds quite generally. In particular, only rather nonrestrictive assumptions about payoff distributions and prior beliefs need to be made. We provide a detailed discussion of the intuition underlying the main result, whereas the formal proof is delineated in the Appendix. Then we discuss the

theoretical implications of our main result and the relations with other sampling explanations of biased judgments. Finally, we discuss how our rational analysis can help explain behavioral regularities previously attributed to biased or motivated information processing.

A Simple Setting

Consider the following task: A decision maker has to choose, in each of 20 periods, between two actions. The actions have two possible outcomes: 0 (failure) and +1 (success). If the decision maker selects the first action (I), she gets +1 with probability p^I or 0 with probability $1 - p^I$. If the decision maker selects the second action (S), she gets +1 with probability p^S or 0 with probability $1 - p^S$. In addition, she observes a realization of the payoff of action I , in spite of the fact that she has not selected it. That is, even if the decision maker does not select I , she can still observe its outcome; however, in this case, she does not receive the payoff from I . We assume that the success probabilities p^I and p^S remain identical throughout all 20 periods and are drawn (at the start of the first period) independently from a uniform distribution ($U(0, 1)$).

In this task, the two actions differ in terms of access to outcome information (see Figure 1): The decision maker has to select S to observe its outcome, but she observes I 's outcome in every period, no matter what. To emphasize this difference, we say that action S has *selective outcome feedback*. Action I is characterized by systematic access to outcome information. In other words, *incidental information* is available whenever action I is not selected. We use the qualifier “incidental” to emphasize the fact that the decision maker does not have to actively select action I to learn about its payoff distribution. In the above scenario, the first action (I) corresponds to interacting with a junior scholar in your department, whose accomplishments you hear about even if you do not interact with him or her. Even if you do not interact with your junior colleague, you obtain incidental information about his or her achievements. The second action (S) corresponds to interacting with a junior scholar from another institution, whose accomplishments you only hear about if you read his or her articles or attend his or her talks.

In what follows, we analyze the beliefs and choices of a risk neutral and rational decision maker who maximizes her expected total payoff. By rational, we mean that (1) she knows that the successive payoffs of the two actions are independent draws from Bernoulli distributions with unknown probabilities of success p^I and p^S ; (2) she knows that p^I and p^S are independent draws from a uniform distribution $U(0, 1)$ —in other words, her priors are correctly specified; (3) she updates her beliefs according to Bayes's

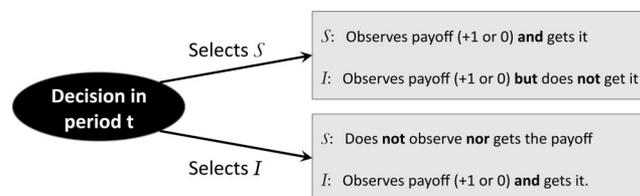


Figure 1. In each period, the decision maker observes the payoff of action I , no matter what action is selected. The decision maker has to select S to observe its payoff.

rule; and (4) she is able to compute the policy that would maximize her total expected payoff over the 20 periods.

The Optimal Policy

To examine how a rational decision maker would behave in this setting, we have to compute the optimal policy. Before we explain how this can be done, it is useful to note that the decision problem analyzed here can be seen as a version of the two-armed bandit problem (Bellman, 1956; Berry & Fristedt, 1985; Gittins, 1989; Gittins & Jones, 1974). In our version, incidental information is systematically available for action I . This differs from the standard two-armed bandit problem, because in that setup outcome feedback is selective for both actions. However, as in a standard two-armed bandit setting, the optimal policy has to solve the tradeoff between exploration and exploitation: To maximize her total payoff, the decision maker has to try to learn which action has the highest quality (here, the highest success probability) while at the same time minimizing the number of times she selects the action with the lowest quality. The optimal policy solves this tradeoff by instructing the decision maker which action to take given her past observations.

There is no simple analytic formula for the optimal policy; it can only be computed numerically by using a computer to find the action that, in any period and given the observed sequence of successes and failures, is the optimal choice. The computations can be simplified, however, by using a standard stochastic dynamic programming procedure called *backward recursion* (Ross, 1983).¹ The result is a “table” that lists optimal choices as a function of the period and the observations made so far. For example, the table instructs the decision maker to select S in the first period.

Implications of the Optimal Policy

To examine how a rational decision maker would behave in the above problem, we used computer simulations to analyze how beliefs and choices evolve over time for a decision maker who follows the optimal policy. Each simulated decision maker starts with uniform priors on the success probabilities of I and S . In each period, the decision maker chooses the action according to the “table” of optimal choices, makes the corresponding observations, and updates her counts of successes and failures of the two actions.

Figure 2 shows how the probability of selecting S changes over time. S is initially selected more often than I ; however, in later periods, S is chosen less often than I . For example, in the last period, S is selected only 45% of the time despite the fact that S is the best action 50% of the time. Thus, the decision maker ends up favoring I .

What about the estimates of the qualities of the two actions? As illustrated by Figure 3, the distribution of estimates of I 's success probability is symmetric, whereas the distribution of the estimates of S 's success probability is skewed toward lower values. Thus, there is a tendency to underestimate the quality of action S , but there is no such tendency for I .

¹ Annotated Matlab code of the programs used for computation of the table of optimal decisions and for the simulations is available in the supplemental materials. Readers interested in learning about how the optimal policy is computed are encouraged to look into it.

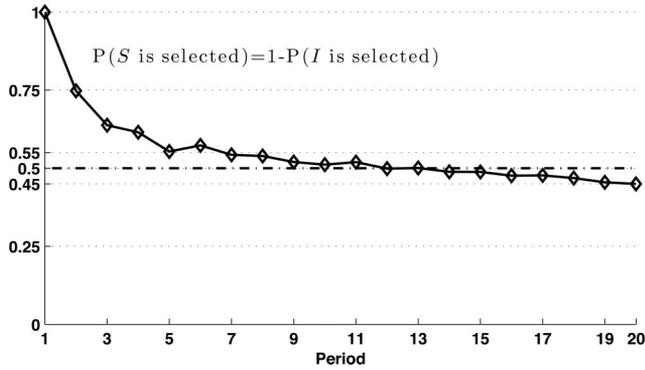


Figure 2. Initially, the action with selective outcome feedback (*S*) is chosen more often than the action with access to incidental information (*I*). However, ultimately there is a reversal in choice probabilities, and *I* is more often selected in the later periods. These calculations are based on 100,000 simulation runs.

Overall, the decision maker ends up choosing *I* more often and is less likely to underestimate its quality, despite the fact that *I* and *S* are each equally likely to be superior and the fact that the decision maker knows this. This tendency to prefer the action with systematic access to incidental information (*I*) can be explained by the adaptive selection of the actions and thus the adaptive sampling of payoff information.

Given the difference in access to outcome information, it is adaptive to select *S* initially. A rational decision maker knows she can only learn about the quality of *S* by selecting it and that she can learn about the value of *I* even if she does not select it. There is therefore some additional information value associated with selecting *S* versus selecting *I*. Because there are only finitely many periods, it is more useful to learn about *S* early in the process. In fact, it is optimal to select *S* in the first period, even if the priors about *S* and *I* are initially identical.

In subsequent periods, whether it is optimal to continue to choose action *S* depends on past outcomes. If the decision maker obtains several failures when she selects *S* in the first few periods, whereas she observes a higher proportion of successes with *I*, she

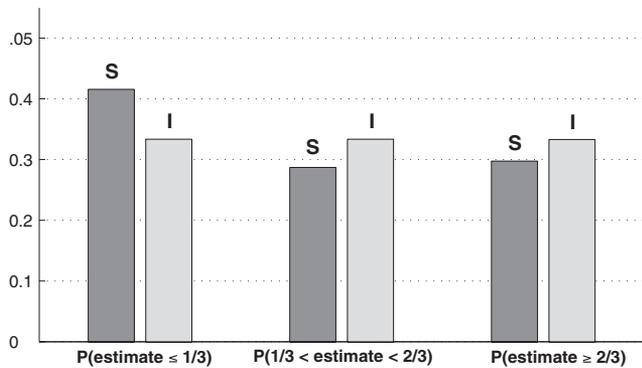


Figure 3. Distributions of the estimates of the qualities of the two actions after Period 20. The distribution of the estimate of *S*'s quality is skewed, whereas the distribution of the estimate of *I*'s quality is symmetric. These calculations are based on 100,000 simulation runs.

might stop selecting *S* so as to exploit her knowledge and maximize her total payoff. By choosing *I*, however, the decision maker will not be able to observe the payoff of *S*. This implies that errors of underestimating the quality of *S* may not be corrected. To explain this, suppose that the decision maker is unlucky with *S* in the sense that she *mistakenly* comes to believe that *S* is inferior to *I* and thus selects *I*. Because she does not observe *S* any longer, she might not be able to discover her mistake.

If, on the contrary, she is lucky with *S* in the first few periods, and mistakenly comes to believe that *S* is superior to *I*, she will keep on selecting *S*. By choosing *S* she will learn about the success probabilities of both actions. As a result, she will most likely discover her mistake before the end of the 20 periods.

Overall, the difference in the probability of correcting mistakes imply that errors of underestimation of the quality of *S* become more likely than errors of overestimation. As Figure 3 shows, there is no such asymmetry in the probabilities of over- or underestimating the quality of *I*. The reason is that the decision maker observes the payoffs of this action in every period.

The asymmetry in errors of over- and underestimation implies a similar asymmetry in beliefs about which action is superior. As illustrated by the graphs of Figure 4, errors of mistakenly believing that *S* is inferior to *I* become more likely than errors of mistakenly believing that *S* is superior to *I*. In fact, at the end of the last period, decision makers are about twice as likely to mistakenly believe that *S* is inferior to *I* than to mistakenly believe that *S* is superior to *I*.

Summary

The simulations of the optimal policy show that it leads to a tendency to underestimate the quality of the action with selective outcome feedback (*S*). There is no such tendency for the estimate of the quality of the action with systematic access to outcome information (*I*).

We assumed that the decision maker was fully aware of the properties of the task (she knew the prior distributions of the success probabilities of the two actions) and updated her estimates according to Bayes's rule. This implies that the belief asymmetry observed in this context cannot be explained by errors in belief

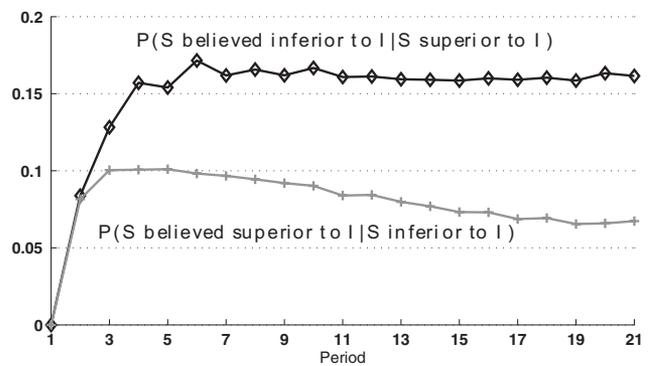


Figure 4. Evolution of the probability of errors in estimates. After a few periods, errors that consist in mistakenly believing that *S* is inferior to *I* become more frequent than errors of mistakenly believing that *S* is superior to *I*. These calculations are based on 100,000 simulation runs.

updating of the quality estimates or incorrect inferences from the samples (i.e., the observations). Furthermore, because we also assumed that, in each period, the decision maker chooses the optimal action with respect to her goal, the belief asymmetry cannot be explained by irrational avoidance of alternatives with poor outcomes.

So, why does the belief asymmetry emerge in that context? The only remaining possibility is that it emerges as an inherent consequence of the interaction between the structure of the environment (an asymmetry in access to information) and the goal of the decision maker (maximizing the total expected payoff), which leads to the avoidance of actions with poor past outcomes. This leads to the acquisition of a biased sample of information about S . In the following section, we demonstrate that this belief asymmetry emerges in a large class of settings.

A General Result on Rational Learning in Infinite Horizon

The above analysis was based on simulations over 20 periods and assumed a particular payoff distribution. In this section, we prove that the basic result holds quite generally. In particular, it is not necessary to assume that the payoffs follow a Bernoulli distribution as we did in the previous section. Furthermore, it is possible to characterize the emergence of the tendency to underestimate the value of actions with selective outcome feedback even when the payoff distributions are not symmetric, or when the priors on the qualities of the two actions differ, and when there are infinitely many periods.

The model of this section is an extension of the two-armed bandit problem similar to that of the previous section. However, here, we assume that the horizon is infinite and that future payoffs are discounted. We first describe the model and then discuss the main characteristics of the optimal policy. Finally, we present our main result about the emergence of the belief asymmetry. The detailed formulation of the optimal policy and the proof of Theorem 1 can be found in the Appendix.

Model: Learning by Doing in Infinite Horizon

In each period $t = 1, 2, \dots$, the decision maker selects one of two actions: S or I . The decision maker observes the payoff of action S only if S is selected, whereas she observes the payoff of action I in every period. Thus, as in the previous section, action S has *selective* outcome feedback, whereas I 's outcome is observed in every period.

We assume that the successive payoffs of each action are iid and drawn from payoff distributions with means u^S and u^I . In what follows, we refer to u^S and u^I as the qualities of the actions. In the special case where the payoffs follow a Bernoulli distribution, as assumed in the model of the previous section, u^S and u^I would be the success probabilities, p^S and p^I . As before, we assume that the decision maker does not know the true qualities of the actions but that she has some priors about these. Let Π^S (resp. Π^I) denote the prior on u^S (resp. u^I). Our results hold for many different assumptions regarding payoff distributions, priors, and belief updating rules (see the Appendix for a precise characterization of what needs to be assumed) including the common cases of (1) normally distributed priors and payoffs with Bayesian updating and (2)

Bernoulli distributed payoffs with priors that follow Beta distributions and Bayesian updating.

The decision maker's objective is to maximize the expected sum of discounted payoffs over all periods,

$$E \left[\sum_{t=1}^{\infty} \delta^{t-1} R_t \right] = \int E_{u^S, u^I} \left[\sum_{t=1}^{\infty} \delta^{t-1} R_t \right] d\Pi^S(u^S) d\Pi^I(u^I),$$

where R_t is the payoff of the action chosen in period t , δ is the discount factor ($\delta < 1$), and E_{u^S, u^I} denotes the conditional expectation given that the true qualities of the actions are fixed. The integral with respect to the prior distributions accounts for the fact that the decision maker does not know the true qualities of the actions.

The Optimal Policy

As in the setting of the previous section, the optimal policy solves the tradeoff between exploration and exploitation: It specifies when the decision maker should select an action believed to have suboptimal quality to learn more about its payoff distribution (i.e., explore) and when she should select the action believed to be the best (i.e., exploit).

In Lemma 3 in the Appendix, we show that the optimal policy is an "index policy." Specifically, in each period, it is possible to define an index for each of the two actions, and the optimal policy selects an action with maximal index. The index of an action depends only on the posterior payoff distribution. The index of action I at the beginning of period t is the best estimate of its quality, or more formally, the mean of its posterior. It is denoted as \hat{u}_t^I . The index of action S is a complicated function of the posterior, the Gittins index, denoted as \hat{v}_t^S (see the Appendix for details on the optimal policy and the Gittins index). It is equal to the sum of two terms: the best estimate of S 's quality, \hat{u}_t^S , and a non-negative term ($\hat{v}_t^S - \hat{u}_t^S$). This second term can be interpreted as the information value (or option value) of selecting S rather than I . There is some information value specifically associated to selecting S because selecting S is the only way to learn more about the payoff distribution of this action. Importantly, this implies it is sometimes optimal to select S even if it is believed to be inferior to I , because there is some potential benefit to learn more about S (e.g., it is possible to learn that, in fact, S is superior to I). Such explorative behavior happens when the estimate of S is lower than the estimate of I but when its index is higher than the estimate of I ($\hat{u}_t^S < \hat{u}_t^I$ but $\hat{v}_t^S > \hat{u}_t^I$).

Beliefs and Choices Under the Optimal Policy: The Emergence of an Asymmetric Pattern

We now formulate our main result: A decision maker who follows the optimal policy will almost always discover when the action with systematic access to outcome information (I) is superior. However, more importantly, such a decision maker sometimes fails to discover when the action with selective outcome feedback (S) is superior. Overall, this implies that the decision maker is likely to favor I . This result holds broadly, and only few assumptions about the payoff and belief distributions are needed (see the Appendix for a precise formulation of these assumptions).

In particular, it holds when the decision maker processes new information rationally, according to Bayes's rule.

Theorem 1. Suppose the quality estimates for S satisfy Conditions (a) to (c) and the quality estimates for I satisfy Conditions (a) to (d) (the conditions are detailed in the Appendix). Suppose, in addition, that the decision maker selects an action in each period by following the optimal policy. Then the following statements are true:

(i) The probability of mistakenly believing that S is superior to I becomes negligible: Suppose u^S and u^I are fixed with $u^S < u^I$. Then

$$\lim_{t \rightarrow \infty} P_{u^S < u^I}(\hat{u}_t^S > \hat{u}_t^I) = 0.$$

(ii) The probability of mistakenly believing that S is inferior to I does not become negligible. Suppose u^S and u^I are fixed with $u^S > u^I$. Then

$$\lim_{t \rightarrow \infty} P_{u^S > u^I}(\hat{u}_t^S < \hat{u}_t^I) > 0.$$

(iii) The probability of mistakenly believing that S is inferior to I becomes higher than the probability of mistakenly believing that S is superior to I . Suppose $P(u^S < u^I) > 0$ and $P(u^S > u^I) > 0$. Then, for t large enough,

$$P(\hat{u}_t^S < \hat{u}_t^I | u^S > u^I) > P(\hat{u}_t^S > \hat{u}_t^I | u^S < u^I).$$

(iv) When both actions are equally likely to be the best, I will more often be believed to be superior to S than inferior to S . Suppose $P(u^S < u^I) = P(u^S > u^I) = 0.5$. Then, for t large enough,

$$P(\hat{u}_t^S < \hat{u}_t^I) > P(\hat{u}_t^S > \hat{u}_t^I).$$

To clarify the scope of the theorem, we first rephrase the Technical Boundary Conditions (a) to (d) as follows:

1. The quality estimate of an action converges toward the true quality of the action when the number of observation of payoffs of that action becomes large.

2. The ordering of the quality estimates (whether S is believed to be superior to I) can always be reversed given enough contradictory evidence.

Taken together, these assumptions essentially say that we consider a decision maker who is at minimum a "good" processor of information: Given enough observations of the successive payoffs of an action, she will develop an accurate estimate of the value of the action. Moreover, the decision maker will always end up taking into account new information, however strong her current beliefs. These conditions are satisfied, in particular, when the decision maker is rational in the sense that she has correctly specified priors (i.e., the true qualities u^I and u^S are random draws from the corresponding prior distributions) and updates her estimates according to Bayes's rule. Our theorem thus implies that selective outcome feedback imposes limits on the knowledge that can be acquired by a rational decision maker. However, the above conditions are broader and are also satisfied in other settings. For example, they do not require the estimates to be updated after every observation.

To better understand what the theorem says, suppose first that S is inferior to I , that is, $u^S < u^I$. Suppose, in addition, that the

decision maker got bad experiences with I . She avoids it and thus selects S . Doing so, she will learn more about both S and I , and she will discover her mistake. This suggests that, in the long-run, the probability of mistakenly believing that S is superior to I becomes negligible, as stated in Part (i). To understand why the decision maker has to discover her mistake, we use a *reductio ad absurdum* argument. Suppose she does not discover her mistake and that she keeps on selecting S in every period. Then Condition 1 above implies that after some time the quality estimate of S becomes close to its true quality: $\hat{u}_t^S \approx u^S$. Also, because the information value of one more selection of S becomes low when the decision maker has made many observations, the index of S becomes close to the estimate of S 's quality: $\hat{v}_t^S \approx \hat{u}_t^S$. These two properties, combined, imply that the index of S becomes close to the true quality of S : $\hat{v}_t^S \approx u^S$. However, Condition 1 also implies that the index of action I becomes close its quality: $\hat{v}_t^I \approx u^I$. These two relations together with the assumption that $u^S < u^I$, imply that, after a large enough number of periods, $\hat{v}_t^S < \hat{u}_t^I$. This means that I is selected. However, this later assertion contradicts the initial assumption that S is always selected. It must be that the initial assumption is wrong. The proofs of Lemma 3 and Theorem 1 in the Appendix develop this intuition into a formal proof.

To understand Part (ii), suppose next that S is superior to I , that is, $u^S > u^I$, but that the decision maker selects S and gets a string of bad outcomes. In this case, the index of S will become quite low. If the index falls below the index of I , ($\hat{v}_t^S < \hat{u}_t^I$), the optimal policy is to avoid S in the next period. If S is in fact superior to I but mistakenly believed to be inferior, this error of underestimation of S 's value will persist, because the decision maker can no longer learn about S .

Part (iii) of the theorem states that, as a result of Parts (i) and (ii), errors of underestimating S will, in the long-run, be more likely than errors of underestimating I . Part (iv) concludes that if it is equally likely that S or I is superior, I is more likely to be believed to be superior in the long-run.

To better understand the role of access to outcome information, it is interesting to note that the asymmetries described by Theorem 1 do not generally emerge when outcome feedback is selective for both actions (as in the two-armed bandit problem) or when incidental information is always available for both actions. Thus, this is the asymmetry of access to outcome information that leads to the emergence of an asymmetry in estimates of the qualities of the actions.

Results similar to those formalized in Theorem 1 hold when there are more than two actions. In this context, even if there is only one action like I (with systematic access to outcome information) and several actions like S (with selective outcome feedback), there is a positive probability that all the type S actions will ultimately be avoided, even if the type I action has the lowest value. In other words, the action with systematic access to incidental payoff information creates a negative externality that will affect all the actions with selective outcome feedback.

Robustness of the Predicted Asymmetry

Our rational analysis is made under specific assumptions, because the structure of the environment needs to be fully specified to compute the optimal choice policy. However, the rational analysis underscores two structural features that lead to the asymmetry

in estimates: asymmetric access to outcome information and a hedonistic goal. The insights brought about by the analysis should thus carry over to other settings that share these features (Oaksford & Chater, 2007).

As an illustration, consider the following modification of the setting analyzed in the second section: the decision maker gets an incidental observation of the outcome of S with probability 10% (instead of 0%) and an incidental observation of the outcome of I with probability 75% (instead of 100%). Computer simulations show that the optimal policy unsurprisingly leads to a belief asymmetry similar to what happened in the basic setup: 47% of the simulated decision makers select S in the last period (Period 20), and about 14% mistakenly believe that I is superior, whereas only 8% mistakenly believe that S is superior (based on 10,000 simulations). This belief asymmetry is not as strong as in the baseline case. This is not surprising, however, because the modified setup is characterized by an asymmetry in access to outcome information that is less strict than in the baseline case.

Similarly, it is possible to change the assumptions made about information processing. In the baseline case, we assumed that the decision maker's memory was perfect and that estimates converged toward the actual quality of an action given sufficient evidence (this is a feature of Bayesian updating). However, intuition suggests that a belief asymmetry similar to that of the baseline case will emerge even if there is some memory decay. Despite the fact that the proof of Theorem 1 does not work in that setting,² computer simulations not reported here show that a similar asymmetry in estimates still emerges.

Our analysis also suggests that many reasonable heuristics will lead to systematic belief and behavior asymmetries like those that emerge under the optimal policy. Consider the baseline case but suppose that the decision maker adopts the following heuristics: select S in Periods 1–5 and then select the action with the highest quality estimate for Periods 5–20. After five periods, the decision maker will have made five observations of both S and I . Although her quality estimates will be fairly accurate, errors are still possible. Furthermore, errors of mistakenly believing that S is inferior to I will be less likely to be corrected than errors of mistakenly believing that S is superior to I —because believing that S is lower than I will lead the decision maker to avoid S and to stop getting additional information about its quality. The asymmetry in estimates will tend to be weaker here than with the optimal policy, however, because there is a form of forced exploration in the first five periods.³

Other realistic situations exist in which the belief and choice asymmetries would be weak or even eliminated under the optimal policy. Suppose there are switching costs. In such setting, decision makers will be unlikely to switch to another alternative after only one or two poor outcomes. Instead, they will tend to stay with the chosen alternative for many periods. Doing so, they will develop a more accurate estimate of the chosen alternative, and the belief asymmetry will likely be weaker than in the baseline case. In settings with non-stationary payoffs distributions, the belief asymmetry will also tend to be weaker than in the baseline case. This is because decision makers may have incentives to return to previously abandoned alternatives. Such behavior could potentially eliminate the bias in the long-run.⁴

Theoretical Implications

On the Possibility to Correct for the Asymmetry in Estimates

The fact that the tendency to underestimate S relative to I emerges even in the context of the optimal policy, and Bayesian updating proves that this tendency is not the consequence of biased information processing. Rather, it is an inherent consequence of the hedonistic goal of the decision maker combined with asymmetric access to information.

A rational decision maker is able to predict, ex-ante, that she is likely to underestimate S relative to I . In other words, we have the surprising result that although most rational decision makers will learn to underestimate S relative to I even when the priors on the qualities of these actions are the same, decision makers have no reason to correct for this tendency *even if they are aware of it*. Stated differently, when the goal is to maximize the expected sum of discounted payoffs, there is *no rational basis* for correcting for this tendency.

The reason for this result is that the decision makers' quality estimates of S and I are *statistically* unbiased estimators. That is, the expected value, across decision makers, of the errors in estimates is equal to zero for S as well as for I .⁵ However, the error distribution for S is skewed: More than 50% of all decision makers underestimate the quality of S . This is not the case for I .⁶

Although there is no rational basis for decision makers to apply a correction to their estimates of S 's value, they could of course reduce the asymmetry in estimates by changing their sampling behavior. For example, a policy with a fixed probability of selecting each alternative in each period would not lead to the belief asymmetries described above. However, it would not allow the decision maker to satisfy her goal to maximize her total expected (discounted) payoff. Even worse, it would not even be adaptive, because information about observed payoffs would not be used to guide future choices. More generally, any alteration of the sampling behavior that deviates from the optimal policy would lead to a lower total expected discounted payoff.

² It uses Lemma 3 in the Appendix, which relies on the assumption that the variance of the distribution of estimates converges to 0 when the number of observations becomes large. With memory decay, the estimates will not converge to a unique value. Instead, they will remain somewhat noisy.

³ A total of 51.4% of the decision makers prefer I ; the probability of mistakenly preferring S is 12.2%, and the probability of mistakenly preferring I is 9.4%. These calculations are based on 1,000,000 simulations.

⁴ These are conjectures; computing the optimal policy when there are switching costs or in non-stationary environments is very difficult, and we have not attempted this.

⁵ More formally, $E(\hat{u}_t^S) = E(u^S)$ and $E(\hat{u}_t^I) = E(u^I)$ for all t . This follows from the optional stopping theorem for martingales (see Billingsley, 1995).

⁶ These claims apply only to settings where the decision maker has correctly specified priors, updates her beliefs according to Bayes's rule, and the priors on the mean payoffs of S and I are the same.

Extension to Other Sampling Explanations of Biased Judgments

Our model shows that systematic judgment errors (here favoring I) can emerge when the decision maker (1) optimally selects actions to achieve her goal, given that she has full knowledge of the structure of the environment including how information is accessed and also correctly specified prior beliefs and (2) rationally processes observed information (using Bayes's rule). These assumptions guarantee that our decision maker is not "naive" with respect to the bias in the sample of information she uses to form her estimates.

In contrast to our active sampling setting, many existing sampling explanations of judgment biases focus on situations where the sample of information available to the decision maker is not actively selected but is determined from prior experiences or from exogenous constraints in the environment (for reviews, see Fiedler, 2000; Fiedler & Juslin, 2006). Will these other sampling explanations continue to be valid even if decision makers are rational (i.e., know how the sample of data they have access to has been generated and process information according to the laws of probability)? It is difficult to say without working out the details of the rational analysis for each specific case.

However, the key to our result is the fact that the distribution of the sample estimates of the value of the action with selective feedback is skewed (across decision makers). In our analysis, this was an outcome of the optimal choice policy given the goal of total payoff maximization. However, a skewed distribution of estimates can also emerge in settings where information is not actively sampled to satisfy a particular goal. For example, Kareev (1995, 2000) noted that sample estimates of correlation (Pearson product-moment correlation coefficients) between two variables tend to be skewed when the sample used for the estimation is small.⁷ An important but largely overseen implication of Kareev's analysis is that even if people were aware of the fact that small samples lead to a tendency to overestimate correlations, they would not be able to correct for this tendency. Although the sample correlation is not always an unbiased estimator of the population correlation, applying the appropriate correction will lead to a statistically unbiased estimator but will not eliminate the skew. In that setting as well, the naivety of the decision maker with respect to the relation between environment and available information sample is not necessary to explain the systematic tendency to flawed judgment.

More generally, people who have a full understanding of how the sample they use to produce estimates has been generated could potentially correct for the bias in their sample estimates—assuming they have the cognitive capabilities to do so. Noting that people generally lack this ability is at the core of research that explains judgment and decision biases from biases in sampled information (Fiedler & Juslin, 2006). However, the distribution of corrected estimates might still remain skewed even after proper correction has been applied. When skewed estimators are used as the basis for judgment, systematic patterns will emerge: Either most people who use a skewed estimator will overestimate the population value of interest, or most people will underestimate it. This is not to say that statistically unbiased but skewed estimators are bad or inappropriate—in some settings, they might not exist any better alternative, as illustrated in the analysis of the previous sections. Rather, the implication is that observing that most people

underestimate a quantity is not enough to conclude that there is flaw in how information is processed, or that people do not properly take into account the relations between environment and available sample of information.

In our setting, the skewed estimates of the quality of action S emerge because the goal of the decision maker is not pure information search (see the discussion of disinterested information search below). In a recent article, Fiedler (2008) has also pointed out the crucial importance of the decision maker's goal on the emergence of sample-based errors in judgments. Fiedler's "ultimate sampling dilemma" refers to the fact that an information sampling strategy that is tailored to answering a certain question might produce a sample that will "distort other judgments for which the sample was not tailored" (p. 190). Fiedler used computer simulations to illustrate how this can happen and produced experimental evidence in support to this point. Our results can be seen as an instance of this ultimate sampling dilemma, but where the dilemma is not between two types of questions that have to be answered but rather between maximizing payoffs and developing accurate estimates.

Disinterested Information Search

In the introduction, we stated that our main result emerges when information search is *interested*. We therefore assumed that the decision maker had a particular hedonistic goal, that of maximizing the expected sum of (discounted) payoffs. To achieve this goal, the decision maker had to balance information search (i.e., exploration) and using this newly acquired knowledge so as to obtain the highest possible payoff (i.e., exploitation). However, what if the decision maker was just interested in knowing which action has the highest expected payoff? In this setting, information search would be disinterested, and the judgment bias in favor of I would not emerge. Thus, in this sense, the emergence of the bias does hinge upon the goal of the decision maker.

To explain this in more detail, suppose the goal of the decision maker is to find out whether S or I is the superior alternative (i.e., to find out whether $u^S > u^I$ or $u^S < u^I$ is the correct hypothesis). In each period, the decision maker can select action I or action S . Optimal data selection in this setting involves choosing, in each period, the action that is most "informative" or "useful" for deciding between the two hypotheses. Researchers have proposed several different ways to formalize these criteria (see Nelson, 2005, for a review); however, before focusing on a specific criterion, it is worth noting that selecting S leads to the acquisition of more information than selecting I , because it leads to an observation of both S and I (because I 's payoff is observed in every period regardless of the decision maker's choice). This suggests that any reasonable sampling rule for that problem should select action S in every period. If S is selected in every period, however, the available information sample will not be biased or asymmetric. Rather, at the end of any period t , the decision maker will have made t observations of the payoff of action S and t observations of the payoff of action I . If the prior distributions of the two alternatives are identical, this implies that the decision maker will be equally likely to believe that S or I is superior.

⁷ Under certain assumptions, the skew tends to be maximal when the sample used to compute the correlation is of about seven observations.

To verify this intuition, we computed the expected information gain (Oaksford & Chater, 1994, 2007) associated with selecting I or S in the setting of the simple model analyzed above. For all combinations of prior beliefs on I and S ,⁸ the expected information gain of selecting action S is higher than the expected information gain of selecting action I . A decision maker that would select, in each period, the alternative with the highest expected information gain would therefore select S in every period. There would thus be no sample bias and no belief asymmetry.

Empirical Implications

In the previous sections, we have emphasized the theoretical implications of our results for the “sampling approach” research program. However, our results also have implications for other areas of research in psychology. More precisely, we believe they offer an alternative explanation of several well-known empirical regularities in impression formation and social influence. Researchers in psychology usually have attributed these regularities to biased information processing or motivated reasoning, but our results suggest that they can also be the outcome of rational behavior and information processing. To develop this claim, we first discuss the psychological phenomena our model helps to explain. Then, we describe the experimental evidence for the mechanism we propose. Finally, we discuss how our rational explanation of these phenomena complements existing explanations that rely on heuristics.

Psychological Phenomena Explained

Our main result shows that even rational decision makers end up favoring alternatives for which information is more accessible. To understand the broader significance of this seemingly technical result, it is important to note that several well-known regularities in social psychology can be explained by invoking systematic differences in access to information. Consider, for instance, the commonly observed phenomenon that people develop more positive opinions of proximate others to whom they are frequently exposed. For example, field studies have shown that college students often have more positive opinions of their roommates than of other students (Festinger, Schachter, & Back, 1950; Marmaros & Sacerdote, 2006; Segal, 1974), and members of an ethnic group often have more positive opinions toward those in their own group than toward members of other groups (Hewstone, Rubin, & Willis, 2002; Levin, van Laar, & Sidanius, 2003).

In psychology, this pattern of belief formation is usually attributed to motivational influences (Cialdini & Goldstein, 2004; Kunda, 1990; Wood, 2000), flawed information processing, or an inherent tendency to like familiar objects (Zajonc, 1968). Considerable experimental support exists for these mechanisms. Nevertheless, as emphasized by Denrell (2005), these experiments typically leave out an important aspect of social reality: Information is often not exogenously given but has to be actively sampled. In addition, access to information is likely to differ systematically between various groups. In particular, you are likely to have access to information about those close to you even if you do not personally interact with them. However, you are less likely to have access to such “incidental” information about distant others. Our model shows that this difference in terms of access to incidental

information has a systematic effect on belief formation even if decision makers are rational. In particular, when individuals have a “hedonistic goal”—when they want to have enjoyable experiences with others—and when proximity is positively correlated with access to “incidental” information, then the tendency to prefer proximate others is a consequence of rational learning. Although our model builds on Denrell’s model of experience sampling, our results go beyond his and show that the effect of information access survives even if decision makers are aware of the sample bias in the data they have access to. In a similar way, our model goes beyond other sampling explanations of in-group bias (e.g., Fiedler, 1996, 2000) in that it shows that the “naivety” assumption is not necessary.

Our model thus offers an alternative, rational explanation of why one would observe that people have more positive opinions about proximate others. It is important to note, however, that our explanation, which relies on a conjunction of a hedonistic goal and an asymmetry in access to outcome information, does not have anything to say about the results of experimental studies where experimenters provide people with the information they have to react to. It is only relevant to field studies and experiments in which people have to actively sample information for some alternatives (see Denrell, 2005, for more detailed discussion).

In addition to in-group bias, our results can also help explain other related psychological phenomena. Consider the well-documented tendency to favor alternatives chosen by many others (see Cialdini & Goldstein, 2004, for a review of the underlying social influence mechanisms). The dominant explanations rely on social influence processes that involve imitative behavior or conformity to norms. However, our model suggests a novel mechanism. The general idea is that the social environment tends to provide additional information about popular alternatives, even if the decision maker does not personally choose them. For example, popular restaurants get reviewed, and thus one can learn about those even if one does not attend them. However, information about new or unpopular venues is harder to access, and one often has to go there to learn about the venue. If the decision maker avoids a popular restaurant following poor experiences, she might still learn about it by reading reviews and learn that it is not that bad. This is less likely to happen for the unpopular restaurant. This asymmetry in access to information can help explain why more popular alternatives are often more positively evaluated. Our rational analysis also leads to distinct predictions: For example, it predicts that the effect will emerge even if the additional information is unbiased. Empirically testing this effect of unbiased information on diffusion and adoption is an interesting avenue for future research.

⁸ In that setting, the prior beliefs on u^I and u^S follow Beta distributions in every period. Let’s consider a specific period t . Let $\hat{u}_t^I \sim \text{Beta}(1 + a_t, 1 + b_t)$ and $\hat{u}_t^S \sim \text{Beta}(1 + a_s, 1 + b_s)$, where a_t and b_t (resp. a_s and b_s) are the numbers of successes and failures observed for action I (resp. S). When there are 20 periods, we have $a_t + b_t \leq 20$ and $a_s + b_s \leq 20$. For all possible combinations of a_t , b_t , a_s , and b_s , the information gain associated to selecting S is superior to the expected information gain associated to selecting I .

Experimental Evidence for the Effect of Incidental Outcome Information

Our main result illustrates the positive effect of access to incidental outcome information on quality estimates (most rational decision makers will come to prefer *I* to *S*). Several experiments have found a similar effect of access to incidental outcome information when the goal of the participants is hedonistic. For example, Fazio, Eiser, and Shook (2004), in an experiment on attitude formation, have shown that giving participants access to information about the outcome of alternatives that were not chosen leads to more positive attitudes. Furthermore, Fetschenhauer and Dunning (2010), in an experiment on trust, have shown that access to incidental outcome information leads to higher trust levels. Experiments on the effect of information about foregone payoffs (i.e., access to incidental outcome information, in the terminology of our article) on choices between uncertain alternatives are also consistent with our model (Erev & Barron, 2005; Grosskopf, Erev, & Yechiam, 2006; Yechiam & Busemeyer, 2006; Yechiam, Stout, Busemeyer, Rock, & Finn, 2005). They found that additional access to information leads to a higher proportion of choices of that alternative.⁹ Although not conclusive, this evidence suggests that access to information can have the effects that our model predicts.

Implications of a Rational Explanation

The above phenomena and experimental results are consistent with our rational model but also with the heuristic models proposed earlier (e.g., Denrell, 2005, 2007). Given this, what is the value of a rational analysis? Like other rational analyses in psychology, our analysis operates at Marr's (1982) computational level of analysis. Our focus is on how a cognitive system faced with the problem of balancing exploration and exploitation should behave when there is an asymmetry in information access. As with other rational analyses, we believe that understanding the optimal solution to this problem helps us better understand why the patterns and behaviors discussed above emerge and persist. This understanding has important normative implications as well as distinct empirical implications.

Normative implications. Knowing that behaviors that appear to be irrational could be the optimal solution to the problem people face has implications for how to change how people behave. Take in-group bias, for example. To eliminate judgment bias, it may not be sufficient to improve people's abilities to interpret information and draw inferences from sample to population characteristics. If in-group bias is a rational response to asymmetric information access, information about distant others needs to be provided and, at the limit, almost "forced upon" those who have little contact with the out-group. A recent field experiment by Shook and Fazio (2008) illustrates how this kind of additional exposure can lead to improvements in attitudes toward the out-group members: The automatically activated racial attitudes of White freshmen that were randomly assigned an African American roommate improved over the course of their first quarter at university. However, the racial attitudes of those who were assigned a White roommate did not change.

Another way to limit judgment biases is to act on the goals of the decision maker. Motivating people to forego their hedonistic

goals in favor of disinterested inquiry should help the formation of more accurate judgments. More generally, noting that judgment biases can be the outcome of rational behavior calls for solutions that alter the task facing the decision maker, the goal, and/or the structure of the environment—rather than how she tries to solve that task.

Parsimony and distinct empirical implications. Our rational analysis explains why even optimal sampling and rational information processing leads to an asymmetry in estimates. It goes beyond existing articles that have demonstrated how learning algorithms based on reinforcement-learning principles can lead to such an asymmetry (Denrell, 2005, 2007; Eiser, Fazio, Stafford, & Prescott, 2003; Eiser, Stafford, & Fazio, 2008). These models make detailed assumptions about the algorithms people use to balance exploration and exploitation and to update their estimates. We make no such assumptions but only assume that the decision maker or cognitive system uses the policy that maximizes expected payoffs. In this sense, the rational analysis is more parsimonious than models based on heuristics.

Moreover, the rational model also leads to predictions that differ from most heuristic models. For example, in contrast to Denrell's (2005) heuristic model, the rational model presented in the second section predicts an initial tendency to select *S* rather than *I* in the first few periods because choosing *S* provides more information that can be used to improve expected payoffs. By adding extra assumptions, heuristic models can be made consistent with this prediction but at the cost of introducing extra parameters.

Conclusion

Sampling explanations of seemingly biased judgments do not have to rely on an assumption of "naivety." It is not necessary to assume that decision makers are not aware of the biases in the samples of information they use to form judgments. Rather, systematic judgment errors can emerge solely because of asymmetries in the information available to the decision maker when her goal is to maximize her expected payoff. More generally, showing that the naivety assumption is not necessary for sampling explanations expands the scope and range of applicability of this class of explanations. Besides, our results show how a rational analysis of a process of belief formation and sequential choices can help explain seemingly biased judgments. Several scholars have recently suggested that rational models of decision making and causal inference can help us understand the goals and functions of cognitive processes (Anderson, 1990; Howes, Lewis, & Vera, 2009; Oaksford & Chater, 1994; Tenenbaum & Griffiths, 2001). Here, we have shown that a rational analysis can also help explain prominent biases in judgment and impression formation. In fact, our model illustrates how a rational analysis of the task facing decision makers can provide a foundation for sampling explanations. What seems to be irrational behavior may in fact be rational

⁹ Moreover, our model predicts that this tendency should be present even for symmetric payoff distributions and not only for the highly skewed distributions often used in experiments. Also, existing experiments compare beliefs and choice across conditions; however, we are not aware of any published experiment that has examined choice and belief formation in an asymmetric bandit setting like the one analyzed in this article.

responses that result from asymmetric, but nevertheless adaptive, search processes.

References

- Anderson, J. R. (1990). *The adaptive character of thought*. Mahwah, NJ: Erlbaum.
- Bellman, R. E. (1956). A problem in the sequential design of experiments. *Sankhya*, *16*, 221–229.
- Berry, D. A., & Fristedt, B. (1985). *Bandit problems: Sequential allocation of experiments*. New York, NY: Chapman Hall.
- Billingsley, P. (1995). *Probability and measure*. New York, NY: Wiley.
- Brezzi, M., & Lai, T. L. (2000). Incomplete learning from endogenous data in dynamic allocation. *Econometrica*, *68*, 1511–1516.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*, 591–621.
- Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. *Psychological Review*, *112*, 951–978.
- Denrell, J. (2007). Adaptive learning and risk taking. *Psychological Review*, *114*, 177–187.
- Denrell, J., & Le Mens, G. (2007). Interdependent sampling and social influence. *Psychological Review*, *114*, 398–422.
- Denrell, J., & Le Mens, G. (2008, August). Illusory correlation as the outcome of experience sampling. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 421–426). Austin, TX: Cognitive Science Society.
- Eiser, R. J., Fazio, R. H., Stafford, T., & Prescott, T. (2003). Connectionist simulation of attitude learning: Asymmetries in the acquisition of positive and negative evaluations. *Personality and Social Psychology Bulletin*, *29*, 1221–1235.
- Eiser, R. J., Stafford, T., & Fazio, R. H. (2008). Expectancy confirmation in attitude learning: A connectionist account. *European Journal of Social Psychology*, *38*, 1023–1032.
- Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review*, *112*, 912–931.
- Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology*, *87*, 293–311.
- Festinger, L., Schachter, S., & Back, K. W. (1950). *Social pressures in informal groups: A study of human factors in housing*. Stanford, CA: Stanford University Press.
- Fetchenhauer, D., & Dunning, D. (2010). Why so cynical? Asymmetric feedback underlies misguided skepticism regarding the trustworthiness of others. *Psychological Science*, *21*, 189–193.
- Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*, 193–214.
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*, 659–676.
- Fiedler, K. (2008). The ultimate sampling dilemma in experience-based decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 186–203.
- Fiedler, K., & Juslin, P. (2006). Taking the interface between mind and environment seriously. In K. Fiedler & P. Juslin (Eds.), *Information sampling and adaptive cognition* (pp. 3–29). Cambridge, England: Cambridge University Press.
- Gittins, J. C. (1979). Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society, Series B (Methodological)*, *41*, 148–177.
- Gittins, J. C. (1989). *Multi-armed bandit allocation indices*. New York, NY: Wiley.
- Gittins, J. C., & Jones, D. M. (1974). A dynamic allocation index for the sequential allocation of experiments. In J. Gani (Ed.), *Progress in statistics* (pp. 241–266). Amsterdam, the Netherlands: North Holland.
- Grosskopf, B., Erev, I., & Yechiam, E. (2006). Foregone with the wind. *International Journal of Game Theory*, *34*, 285–302.
- Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choices. *Psychological Science*, *15*, 534–539.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, *53*, 575–604.
- Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, *116*, 717–751.
- Juslin, P., Winman, A., & Hansson, P. (2007). The naive intuitive statistician: A naive sampling model of intuitive confidence intervals. *Psychological Review*, *114*, 678–703.
- Kareev, Y. (1995). Through a narrow window: Working memory capacity and the detection of covariation. *Cognition*, *56*, 263–269.
- Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlation. *Psychological Review*, *107*, 397–402.
- Kathakis, M. N., & Veinott, A. F., Jr. (1987). The multi-armed bandit problem: Decomposition and computation. *Mathematics of Operations Research*, *12*, 262–268.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.
- Levin, S., van Laar, C., & Sidanius, J. (2003). The effects of ingroup and outgroup friendships on ethnic attitudes in college: A longitudinal study. *Group Processes & Intergroup Relations*, *6*, 76–92.
- Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York, NY: Wiley.
- March, J. G. (1996). Learning to be risk averse. *Psychological Review*, *103*, 309–319.
- Marmaros, D., & Sacerdote, B. (2006). How do friendships form? *The Quarterly Journal of Economics*, *121*, 79–119.
- Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.
- Nelson, J. D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, *112*, 979–999.
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*, 608–630.
- Oaksford, M., & Chater, N. (1998). *Rational models of cognition*. Oxford, England: Oxford University Press.
- Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, England: Oxford University Press.
- Ross, S. (1983). *Introduction to Stochastic dynamic programming: Probability and mathematical*. Orlando, FL: Academic Press.
- Segal, M. W. (1974). Alphabet and attraction: An unobtrusive measure of the effect of propinquity in a field setting. *Journal of Personality and Social Psychology*, *30*, 654–657.
- Shook, N. J., & Fazio, R. H. (2008). Interracial roommate relationships: An experimental field test of the contact hypothesis. *Psychological Science*, *19*, 717–723.
- Simon, H. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*, 99–118.
- Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138.
- Tenenbaum, J., & Griffiths, T. (2001). The rational basis of representativeness. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 1036–1041). Edinburgh, Scotland: Cognitive Science Society.
- Tversky, A., & Kahneman, D. (1974, September 27). Judgment under uncertainty: Heuristics and biases. *Science*, *27*, 1124–1131.
- Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Review of Psychology*, *51*, 539–570.
- Yechiam, E., & Busemeyer, J. R. (2006). The effect of foregone payoffs on

underweighting small probability events. *Journal of Behavioral Decision Making*, 19, 1–16.

Yechiam, E., Stout, J. R., Busemeyer, J., Rock, S., & Finn, P. (2005). Individual differences in the response to forgone payoffs: An examina-

tion of high functioning drug abusers. *Journal of Behavioral Decision Making*, 18, 97–110.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure [Monograph]. *Journal of Personality and Social Psychology*, 9, 1–27.

Appendix

Emergence of Belief Asymmetries Under Optimal Experimentation

The model is a version of the discounted two-armed bandit problem, and some of the standard results for that problem can be extended to that setting. We first characterize the optimal policy for the discounted two-armed bandit with asymmetric outcome feedback, and then we use this characterization to prove Theorem 1.

Optimal Policy of Experimentation

The optimal solution to the standard discounted multi-armed bandit problem is an index policy that chooses, at each stage, an action with maximal Gittins index (Gittins, 1979, 1989; Gittins & Jones, 1974). Existing proof of the optimality of the Gittins index policy in the standard discounted multi-armed bandit setting does not apply here because the two alternatives are not independent: When the decision maker selects action S , she also gets some information about the payoff distribution of action I . Here, we show that, in spite of this non-independence of the alternatives, the optimal policy for the sequential choice between actions S and I is also an index policy. The index of action S is its Gittins index, whereas the index of action I is its subjective expected reward.

Let $n^i(t)$ be the number of times the outcome action i has been observed prior to period t , and Π_n^i is the posterior on u^i when i has been selected n times ($i \in \{S, I\}$). Note that $n^I(t) = t - 1$ because I 's outcome is observed in every period. The subjective expected payoff of action i at the beginning of period t is denoted by $\hat{u}_{i,t} = \int E_{u^i} R_t^i d\Pi_{n^i(t)}^i(u^i)$, where R_t^i is the outcome of action i in period t , $\Pi_{n^i(t)}^i$ is the posterior on u^i given the information available at the beginning of period t (i has indeed been selected $n^i(t)$ times before the beginning of period t). We now formulate the optimal policy:

Lemma 2. Suppose that for all t , $\int (\hat{u}_t^S)^2 d\Pi_{n^S(t)}^S < \infty$. A sufficient condition for a policy to be optimal is that, in any period, it selects an action with maximal index. The index for S at decision time t is its Gittins index:

$$v(\Pi_{n^S(t)}^S) = \sup_{\tau \geq 1} \left\{ \frac{\int E_{u^S} \left[\sum_{k=0}^{\tau-1} \delta^k R_{t+k}^S \right] d\Pi_{n^S(t)}^S(u^S)}{\int E_{u^S} \left[\sum_{k=0}^{\tau-1} \delta^k \right] d\Pi_{n^S(t)}^S(u^S)} \right\},$$

where the sup is over all stopping times $\tau \geq 1$ defined on $(R_t^S, R_{t+1}^S, \dots)$.

The index for I is \hat{u}_t^I .

Proof. The problem is a stationary Markov decision problem where a state is completely defined by the number of selections of action S and the set of all past payoffs obtained for S and I . The state variables uniquely define the information available to the decision maker at a given decision time t . As in Katchikis and Veinott (1987), the proof of the optimality of the index policy is based on policy improvement.

Let π^* be a policy satisfying the hypotheses of the lemma. Let π_0 be an arbitrary policy and π_1 be a policy that is consistent with π^* in the first period and with π_0 afterwards. Let V_π be the value of an arbitrary policy π . Let's show that $V_{\pi_1} \geq V_{\pi_0}$. The optimality of π^* will follow by induction on $t \geq 1$.

At the beginning of period 1, the index of action S is $v(\Pi_0^S)$, whereas the index of action I is $\hat{u}_{1,1}$. We first suppose that $v(\Pi_0^S) > \hat{u}_{1,1}$. If π_0 chooses S in the first period, then $V_{\pi_1} = V_{\pi_0}$. Suppose that, on the contrary, π_0 chooses I in the first period. Let a third action, R , be such as R has a known and constant reward G with $v(\Pi_0^S) > G > \hat{u}_{1,1}$. We assume that selecting R entails an observation of I 's reward. Let π_R be a policy that chooses R in the first period and that is consistent with π_0 afterwards. The policies π_R and π_0 only differ in terms of the reward obtained in the first period. The distribution of the states that the two policies result in is the same because they have identical information acquisition schemes. Because the reward in Period 1 is higher for π_R , we have that $V_{\pi_R} \geq V_{\pi_0}$.

(Appendix continues)

We also have $V_{\pi_1} \geq V_{\pi_R}$. To see why, consider the following two armed bandit problems derived from the choice problem facing the decision maker in Period 1. In each period, the decision maker has to choose between two alternatives, R' and S' , that differ from R and S only by the fact that their selections do not imply any acquisition of information about I . Information about I is irrelevant to the choice between S and R and to the choice between S' and R' . These two decision problems are therefore equivalent and S (resp. R) is preferred to R (resp. S) whenever S' (resp. R') is preferred to R' (resp. S'). The optimal choice between S' and R' is defined by the Gittins index. The Gittins index for S' is $v(\Pi_0^S)$, and the Gittins index for R' is G . Because $v(\Pi_0^S) > G$, S' is preferred to R' . It follows that, in the original decision problem, S is preferred to R and that $V_{\pi_1} \geq V_{\pi_R}$. This implies $V_{\pi_1} \geq V_{\pi_0}$.

If, at the beginning of period 1, $v(\Pi_0^S) < \hat{u}_{I,1}$, we also have $V_{\pi_1} \geq V_{\pi_0}$ (by a similar construction).

Also, if $v(\Pi_0^S) = \hat{u}_{I,1}$, then $V_{\pi_1} = V_{\pi_0}$ because S and I are equally attractive. In summary, $V_{\pi_1} - V_{\pi_0} \geq 0$.

Using the above reasoning, it follows by induction on $t \geq 1$ that there exists π_t that agrees with π^* up until t and that has the property that $V_{\pi_t} - V_{\pi_{t-1}} \geq 0$. So, $V_{\pi_t} \geq V_{\pi_0}$. This implies $V_{\pi^*} = \lim_{t \rightarrow \infty} V_{\pi_t} \geq V_{\pi_0}$. In conclusion, π^* is an optimal policy.

Asymmetric and Incomplete Learning

Now that we have specified the optimal policy for the discounted bandit problem with asymmetric outcome feedback, it is possible to demonstrate the emergence of the belief asymmetry. We first introduce some notations, and some regularity conditions that are necessary for the results to hold.

Notations. Let $\hat{u}_{(n)}^i$ be the mean and $\sigma_{(n)}^i$ be the standard deviation of Π_n^i , the posterior of u^i when action i has been observed n times. Note that, at the beginning of period t , $\hat{u}_t^S = \hat{u}_{S,(n^S(t))}$ and $\hat{u}_t^I = \hat{u}_{I,(t-1)}$. Γ is defined as the common space of possible values for u^i , $i \in \{S, I\}$. We assume that u^i follows a continuous prior distribution, Π_0^i , over Γ . Let l_Γ denote the infimum of the expected payoffs: $l_\Gamma = \inf_{u^i \in \Gamma} (u^i)$.

Regularity conditions. For an arbitrary event A , let $P_u(A)$ denote the conditional probability that A occurs given the value of u . Consider the following four conditions:

(a) $\int (\hat{u}_{(n)}^i)^2 d\Pi_n^i < \infty$ for all n ;

(b) $P_{u^i}(\lim_{n \rightarrow \infty} \hat{u}_{(n)}^i = u^i) = 1$ and $P_{u^i}(\lim_{n \rightarrow \infty} \sigma_{(n)}^i = 0) = 1$;

(c) $P_{u^i}(\hat{u}_{(n)}^i > l_\Gamma) = 1$ for all n ;

(d) For any $\lambda > l_\Gamma$, there exists $m = m(u, \lambda)$, such that

$$P_{u^i}(\hat{u}_{(m)}^i < \lambda) > 0.$$

Note that these conditions are not particularly restrictive as they are met by the following typical settings: (1) prior beliefs about u^i following a Beta distribution and payoffs following a Bernoulli distribution; (2) prior beliefs about u^i following a Normal distribution and payoffs following a Normal distribution with unknown mean u^i and known variance σ_0^2 . Those conditions are also met when u^i is known. The results exposed below thus still hold when the true quality of action I , u^I , is known.

Proof of the emergence of choice asymmetries. We first demonstrate that when the value of I is superior, the optimal policy will converge to that action with probability 1. In contrast, when the value of S is superior, the optimal policy will still converge to I with positive probability. The proof makes use of the above characterization of the optimal policy in terms of the Gittins index and of a result about incomplete learning in standard discounted bandit settings due to Brezzi and Lai (2000).

Lemma 3. Suppose u^S and u^I are fixed.

(i) If $u^S < u^I$, S will be chosen a finite number of times with probability 1.

(ii) If $u^S > u^I$, and $u^I > l_\Gamma$, S will be chosen a finite number of times with positive probability.

Proof. (i) Suppose that $u^S < u^I$. Let $N_\infty(i)$ denote the total number of times action i is selected. Suppose that S is chosen infinitely many times ($N_\infty(S) = \infty$). Conditions (a) and (b) applied to action S imply $v(\Pi_{n^S(t)}^S) \rightarrow u^S$ almost surely (Brezzi & Lai, 2000). Also, by assumption, we have $\hat{u}_t^I \rightarrow u^I$ almost surely. Because

$u^S < u^I$, this implies that, almost surely, there exists t_0 such as if $t \geq t_0$, $v(\Pi_{n^S(t)}^S) < \hat{u}_t^I$. When this occurs, I rather than S is selected. S is therefore almost surely chosen a finite number of times. This implies that the event $\{N_\infty(S) = \infty\}$ cannot occur with positive probability, which concludes the proof.

(ii) The result follows from an easy adaptation of the proof of Theorem 2ii in Brezzi and Lai (2000) to the context of this lemma. We use the fact that the estimate of u^I converges toward the true value while there is a positive probability that the estimate of u^S becomes so low that action S is abandoned and never chosen again.

Proof of Theorem 1

Proof. This asymmetry in the convergence of choices translates into the belief asymmetries formulated in Theorem 1.

(Appendix continues)

(i) By Condition (a) and Theorem 1 in Brezzi and Lai (2000), $\nu(\Pi_{\pi^S(t)}^S) \geq \hat{u}_t^S$ for all t . Thus, we have $P_{u^S < u^l}(\hat{u}_t^S > \hat{u}_t^l) \leq P_{u^S < u^l}(\nu(\Pi_{\pi^S(t)}^S) > \hat{u}_t^l) \leq P_{u^S < u^l}(C_t = S)$, where C_t denotes the action that is chosen in period t .

Lemma 3i implies $\lim_{t \rightarrow \infty} P_{u^S < u^l}(C_t = S) = 0$; (i) follows immediately.

(ii) Note first that

$$P_{u^S > u^l}(\hat{u}_t^S < \hat{u}_t^l) \geq P_{u^S > u^l}(\hat{u}_t^S < \hat{u}_t^l | N_\infty(S) < \infty) P_{u^S > u^l}(N_\infty(S) < \infty).$$

Let us show that $\lim_{t \rightarrow \infty} P_{u^S > u^l}(\hat{u}_t^S < \hat{u}_t^l | N_\infty(S) < \infty) = 1$. Suppose $N_\infty(S) < \infty$. Almost surely, there exists t_0 such as if $t \geq t_0$, $\nu(\Pi_{\pi^S(t)}^S) = \nu(\Pi_{\pi^S(t_0)}^S) < \hat{u}_t^l$. According to Brezzi and Lai (2000), Theorem 1, we also have $\hat{u}_t^S \leq \nu(\Pi_{\pi^S(t_0)}^S)$ for all t . Thus, with probability 1, there exists t_0 such as if $t \geq t_0$, $\hat{u}_t^S < \hat{u}_t^l$. This concludes this stage of the proof.

Thus, we have $\lim_{t \rightarrow \infty} P_{u^S > u^l}(\hat{u}_t^S < \hat{u}_t^l) \geq P_{u^S > u^l}(N_\infty(S) < \infty) > 0$ where the last inequality is a direct consequence of Lemma 3ii.

(iii) Note that (iii) follows immediately from (i) and (ii) and by application of Lebesgue's dominated convergence theorem.

(iv) We have

$$\begin{aligned} P(\hat{u}_t^S < \hat{u}_t^l) &= P(\hat{u}_t^S < \hat{u}_t^l | u^S > u^l) P(u^S > u^l) \\ &\quad + P(\hat{u}_t^S < \hat{u}_t^l | u^S < u^l) P(u^S < u^l) \\ &= 0.5 P(\hat{u}_t^S < \hat{u}_t^l | u^S > u^l) + 0.5 P(\hat{u}_t^S < \hat{u}_t^l | u^S < u^l) \\ &= 0.5 P(\hat{u}_t^S < \hat{u}_t^l | u^S > u^l) + 0.5 [1 - P(\hat{u}_t^S > \hat{u}_t^l | u^S < u^l) \\ &\quad - P(\hat{u}_t^S = \hat{u}_t^l | u^S < u^l)] \\ &= 0.5 + 0.5 [P(\hat{u}_t^S < \hat{u}_t^l | u^S > u^l) - P(\hat{u}_t^S > \hat{u}_t^l | u^S < u^l) \\ &\quad - P(\hat{u}_t^S = \hat{u}_t^l | u^S < u^l)] \end{aligned}$$

Also,

$$\begin{aligned} P(\hat{u}_t^S > \hat{u}_t^l) &= 0.5 P(\hat{u}_t^S > \hat{u}_t^l | u^S > u^l) + 0.5 P(\hat{u}_t^S > \hat{u}_t^l | u^S < u^l) \\ &= 0.5 [1 - P(\hat{u}_t^S < \hat{u}_t^l | u^S > u^l) - P(\hat{u}_t^S = \hat{u}_t^l | u^S > u^l)] \\ &\quad + 0.5 P(\hat{u}_t^S > \hat{u}_t^l | u^S < u^l) \\ &= 0.5 - 0.5 [P(\hat{u}_t^S < \hat{u}_t^l | u^S > u^l) - P(\hat{u}_t^S > \hat{u}_t^l | u^S < u^l) \\ &\quad - 0.5 P(\hat{u}_t^S = \hat{u}_t^l | u^S > u^l)] \end{aligned}$$

Then,

$$\begin{aligned} P(\hat{u}_t^S < \hat{u}_t^l) - P(\hat{u}_t^S > \hat{u}_t^l) &= P(\hat{u}_t^S < \hat{u}_t^l | u^S > u^l) - P(\hat{u}_t^S > \hat{u}_t^l | u^S \\ &\quad < u^l) - 0.5 (P(\hat{u}_t^S = \hat{u}_t^l | u^S < u^l) - P(\hat{u}_t^S = \hat{u}_t^l | u^S > u^l)) \end{aligned}$$

Note that $\lim_{t \rightarrow \infty} P(\hat{u}_t^S = \hat{u}_t^l | u^S < u^l) = 0$. To see why, fix u^S and u^l with $u^S < u^l$. S is chosen a finite number of times with probability 1. When this happens, with probability 1, there exists t_0 such as if $t \geq t_0$, $\hat{u}_t^S < \hat{u}_t^l$ (see proof of (ii) above). Thus, $\lim_{t \rightarrow \infty} P_{u^S < u^l}(\hat{u}_t^S = \hat{u}_t^l) = 0$. Lebesgue's dominated convergence theorem implies that $P(\hat{u}_t^S = \hat{u}_t^l | u^S < u^l)$ converges toward 0 as well.

Similarly $\lim_{t \rightarrow \infty} P(\hat{u}_t^S = \hat{u}_t^l | u^S > u^l) = 0$. To see why, fix u^S and u^l with $u^S > u^l$. In that case, it is possible that S is chosen an infinite number of times, or only finitely many times. When S is chosen an infinite number of times, $\lim_{t \rightarrow \infty} \hat{u}_t^S = u^S$ and $\lim_{t \rightarrow \infty} \hat{u}_t^l = u^l$. Thus, for t large enough, $\hat{u}_t^S > \hat{u}_t^l$. When S is chosen a finite number of times, with probability 1, there exists t_0 such as if $t \geq t_0$, $\hat{u}_t^S \leq \hat{v}_t^S < \hat{u}_t^l$. Hence, $\lim_{t \rightarrow \infty} P_{u^S < u^l}(\hat{u}_t^S = \hat{u}_t^l) = 0$. As above, this convergence result still holds after integration over $\{u^S, u^l | u^S > u^l\}$, which proves that $\lim_{t \rightarrow \infty} P(\hat{u}_t^S = \hat{u}_t^l | u^S > u^l) = 0$.

Finally, i and ii imply that $\lim_{t \rightarrow \infty} P(\hat{u}_t^S < \hat{u}_t^l) - P(\hat{u}_t^S > \hat{u}_t^l) = P(N_\infty(S) < \infty | u^S > u^l) > 0$. The conclusion follows immediately.

Received July 29, 2010

Revision received January 25, 2011

Accepted January 25, 2011 ■

Showcase your work in APA's newest database.



Make your tests available to other researchers and students; get wider recognition for your work.

"PsychTESTS is going to be an outstanding resource for psychology," said Ronald F. Levant, PhD. "I was among the first to provide some of my tests and was happy to do so. They will be available for others to use—and will relieve me of the administrative tasks of providing them to individuals."

Visit <http://www.apa.org/pubs/databases/psyc-tests/call-for-tests.aspx> to learn more about PsychTESTS and how you can participate.

Questions? Call 1-800-374-2722 or write to tests@apa.org.

Not since PsycARTICLES has a database been so eagerly anticipated!