Bayesian Inference as Probability Transfer Across Sample Spaces

Jun Zhang and Zifu Shi

Online First Publication, May 20, 2019. http://dx.doi.org/10.1037/dec0000108

American Psychological Association

Decision

© 2019 American Psychological Association
2325-9965/19/$12.00

http://dx.doi.org/10.1037/dec0000108

Bayesian Inference as Probability Transfer Across Sample Spaces

Jun Zhang
The University of Michigan Ann Arbor

Zifu Shi
Hunan Normal University

Sample space in probability theory refers to the set of all possible outcomes of an experiment or all possible values of a random variable. Probability measures are defined with respect to an associated sample space. When there is more than one random variable (and hence more than one sample space) involved, interaction of these sample spaces can give rise to subtle yet important issues in both classical and modern contextual probability theories and their applications to cognitive modeling. Here we investigate Bayesian reasoning in humans from the perspective of transportation of probability measures across two different sample spaces: namely, a hypothesis space and an evidence (or data) space. Taking Bayesian inference as a task of constructing probability measures, we advance the Probability Transfer Theory (PTT) in which we postulate that probability transfer (transportation) between hypothesis and evidence spaces is susceptible to the structural similarity between these 2 sample spaces. We report experiments on Bayesian reasoning using the commonly used Mammography Problem, the Lottery Problem, as well as three versions of a Factory Problem, all with the same numerical values of probability but varying degrees of structural similarity between the hypothesis space and the evidence space. We analyze the pattern of overestimation and underestimation of Bayesian posterior probability through receiver operating characteristic (or ROC) curve analysis (adapted from the Signal Detection Theory). We find empirical evidence in support of our PTT model of contextual influence on Bayesian reasoning as probability transport.

Keywords: Bayesian reasoning, base-rate neglect, probability transport, the Mammography Problem, Signal Detection Theory

Bayesian inference plays a fundamental role in reasoning about uncertainty. It is a methodology as well as a philosophy for combining two sources of information to derive an *a posteriori* estimate of the probability that a hypothesis is true (or false): (a) the prior knowledge about the truthfulness of the hypothesis and all of its alternatives, and (b) the likelihood of the evidence gathered in support of each hypothesis. This prior-plus-evidence approach to uncertainty reasoning is a powerful computational tool for probabilistic inference and modeling and captures the dynamic evolution of probabilistic beliefs by a normative (i.e., rational) decision maker.

In the Bayesian framework, one deals with two spaces of probability measures: the space Ω_{Hypo} of hypotheses and the space Ω_{Data} of data or evidence. For simplicity, let us consider only the binary case: two mutually exclusive hypotheses, H versus ~H (read as "not H"), with two possible data/evidence, D versus ~D (read as "not D"). That is to say, there are two elements in the hypothesis space $\Omega_{\text{Hypo}} = \{\text{H}, \sim\text{H}\}$, and

Jun Zhang, Department of Psychology and Department of Mathematics, The University of Michigan; Zifu Shi, Cognition and Human Behavior Key Laboratory of Hunan Province, Hunan Normal University.

This work is conducted while Zifu Shi was an exchange visitor at the University of Michigan during 2014–2015 under the support of China Scholarship Council and the National Social Science Foundation of China (Grant BBA160044 to Zifu Shi). Research is also supported in part by ARO Grant W911NF 12–1-0163 and AFOSR Grant FA9550-13–1-0025 awarded to Jun Zhang.

Correspondence concerning this article should be addressed to Jun Zhang, Department of Psychology, The University of Michigan, 530 Church Street, Ann Arbor, MI 48109. E-mail: junz@umich.edu

two elements in the evidence space $\Omega_{\text{Data}} = \{D, \sim D\}$. As an example, in the so-called Mammography Problem (adapted from Edwards, 1968, to be discussed later), H is "breast cancer" while ~H is "no breast cancer", D is "positive test result" while ~D is "negative test result".

Note that both Ω_{Hypo} and Ω_{Data} are called "sample space," which, by definition, is a space on which a probability measure can be defined. So we can speak of:

1. the probability P(H) that hypothesis H (as opposed to its alternative ~H) is true;
2. the probability P(D) of obtaining data/evidence D (as opposed to obtaining its complement ~D);
3. the conditional probabilities P(H|D), which is interpreted as the "posterior" probability that the hypothesis is true after obtaining data D; and
4. the conditional probabilities P(D|H), which is interpreted as the "likelihood" of obtaining data D if hypothesis H were to be true.

The 2 × 2 combination of elements of the hypothesis space and the data space leads to a cross-table of all possible outcomes; in the language of Signal Detection Theory, the entries of this table can be labeled as *hit*, *false alarm*, *miss*, and *correct rejection* outcomes, see Table 1.

Here a, b, c, d denotes the frequencies of occurrence of the corresponding outcome of this 2 × 2 table. Viewed this way, the Bayes formula

$$P(H|D) = \frac{P(D|H) \times P(H)}{P(D|H) \times P(H) + P(D|\sim H) \times P(\sim H)}$$

is nothing but a simple statement of the following identity (see Mandel, 2014):

Table 1
Possible Outcomes With Natural Frequency

$\Omega_{\text{Hypo}} \backslash \Omega_{\text{Data}}$	D	~D
H	Hit *a*	Miss *b*
~H	False alarm *c*	Correct rejection *d*

Note. Ω_{Data}: data space, containing D and ~D. Ω_{Hypo}: hypothesis space, containing H and ~H.

$$\frac{a}{a+c} = \frac{\frac{a}{a+b} \times \frac{a+b}{a+b+c+d}}{\frac{a}{a+b} \times \frac{a+b}{a+b+c+d} + \frac{c}{c+d} \times \frac{c+d}{a+b+c+d}} \quad (1)$$

where the posterior probability is

$$P(H|D) = \frac{a}{a+c}$$

and the hit rate $P(D|H)$, false-alarm rate $P(D|\sim H)$, and prior probabilities $P(H)$ and $P(\sim H)$ are given as

$$P(D|H) = \frac{a}{a+b}; \qquad P(D|\sim H) = \frac{c}{c+d};$$
$$P(H) = \frac{a+b}{a+b+c+d}; \quad P(\sim H) = \frac{c+d}{a+b+c+d}$$

This is the "frequentist" (i.e., objective) interpretation of Bayesianism.

Bayesian Reasoning in Humans and Base-Rate Neglect

Human decision making under uncertainty in the spirit of Bayesian reasoning prevails in our daily life. Common scenarios include (a) medical diagnoses, for example, in an HIV or cancer test, physicians have to understand and to explain what a positive test result really means (Eddy, 1982; Ellis, Cokely, Ghazal, & Garcia-Retamero, 2014; Gigerenzer & Hoffrage, 1995); (b) legal settings, for example, juries must consider a body of evidence that bears upon the innocence or guilt of the defendant (Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000); (c) scientific debates about whether experimental results are consistent with one theory versus another, where people may revise their opinions in the face of new information, and so forth. Edwards and his colleagues (Edwards, 1968; Phillips & Edwards, 1966) empirically tested whether human uncertainty inference follows Bayes's theorem and concluded that probability inference in humans was roughly consistent with Bayesian calculations, although on the "conservative" side. However, Kahneman and Tversky (1972, p. 450), in their seminal investigation of decision heuristics and

biases, revealed that people systematically neglect base rates in Bayesian inference problems—they argued for a kind of "representativeness heuristic" to explain base-rate neglect. Base-rate neglect has since been one of the most consistently found biases in Bayesian reasoning in humans, both in laboratory settings (see recent review of Brase & Hill, 2015) and in real-life medical diagnostic settings (e.g., Brase, Fiddick, & Harries, 2006).

To pin down cognitive mechanism for base-rate neglect in Bayesian inference, several researchers (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995) investigated more natural formats for probabilistic reasoning, that is, using natural frequencies as actually experienced events rather than probabilities or percentages. An advantage was found for the natural frequency format, leading to the notion of "ecological rationality" by Gigerenzer. Mellers and McGraw (1999) agreed with Gigerenzer and Hoffrage that frequency format can improve Bayesian reasoning but found that the beneficial effect of frequency formats was just for the rare event—if all events are common, then the advantage of natural frequencies is reduced (as in Gigerenzer & Hoffrage, 1995). Mellers and McGraw attributed such improvement to the use of mental models that involve elements of nested sets. This "nested-sets hypothesis" (Barbey & Sloman, 2007; Girotto & Gonzalez, 2001; Sloman, Over, Slovak, & Stibel, 2003; Yamagishi, 2003) argues that uncertainty reasoning can be improved irrespective of numerical format, as long as the "nested sets," namely, subsets relative to larger sets in the task structure, is visualized. Johnson-Laird (1999) applied the mental model framework of reasoning to argue that individuals who are unfamiliar with probability calculation can nevertheless infer probabilities of events in an extensional way through constructing mental models—each model represents an equiprobable alternative unless individuals have beliefs to the contrary, in which case some models will have higher probabilities than others. Other researchers even went so far as to invoke two different cognitive systems ("dual-process theory") to explain the role of information format (Kahneman & Frederick, 2005; Sloman, 1996), with the resulting theory supporting the nested-sets hypothesis (Barbey & Sloman, 2007).

Aside from presentation format, researchers have also investigated how context (cover story) affected the performance in Bayesian reasoning. Siegrist and Keller (2011) showed that participants were twice as likely to solve the Bayesian reasoning problem in a social context compared with a problem with medical context. Binder, Krauss, and Bruckmaier (2015) found better performance in almost every version of the economics problem (30% correct inferences, averaged across format of information and visualization) compared to the respective versions of the Mammography Problem (16%), though these two studies used different probability numbers. In an earlier work involving one of the present authors, Tang and Shi (2011) manipulated the levels of the base rate (high/low) and the hit rate (high/moderate/low) in a story using hepatitis and lottery context, with results showing that both the base rate and the hit rate played an important role in Bayesian reasoning.

In addition to issues of presentation format and problem context, there is also evidence of anchoring bias in Bayesian reasoning. Gigerenzer and Hoffrage (1995) adopted 15 different problems, with varying base rates ranging from 0.005% to 80% (i.e., from very rare to very common occurrences). In their Study 2, using a write-aloud protocol in which one of their participants (Rüdiger) explicitly wrote down his reasoning process, they found that this participant focused on p(D|H), remarking that "a child of an infected mother is at such high risk (40%)." It appeared that the participant simply adjusted P(D|H) by 5%, resulting in an estimate of the posterior probability of severe prenatal damage at 35% – this is what Gigerenze and Hoffrage called "adjusted Fisherian" heuristics. Shi and Zhang (2009, 2011) analyzed participates' oral reports and online reasoning process (narrating reasons) to pin-point such anchoring bias as distinct from base-rate neglect.

Our current study continues the long trail of investigations into the source of deviation ("error") from normative prescriptions of Bayesian reasoning. Synthesizing extant theories, in particular the "mental model," we propose a theoretical framework that treats uncertainty inference as probability transfer between two sample spaces, whereas the transfer might be affected by context, in particular structural similarities or relatedness between the concrete sample spaces

involved. This so-called Probability Transfer Theory (PTT) will be used to derive an individual measure that describes sampling and transfer. The investigation of biases in Bayesian reasoning in terms of sampling process is not new. For instance, Kleiter (1994) and Fiedler, Brinkmann, Betsch, and Wild (2000) investigated the advantage of frequency format in terms of sampling processes that mediate probability judgments. Our framework, with supporting evidence, will attribute the bias in Bayesian inference to the underlying cognitive processes involved in sampling and transporting uncertainty from one context (sample space) to another in a context-dependent way.

Probability Transfer Across Sample Spaces

Bayesian inference involves two sample spaces, the hypothesis space denoted Ω_{Hypo} and evidence or data space denoted Ω_{Data}. The conditional probabilities, though, have drastically different interpretations, with P(H|D) interpreted as posterior probability (a derived quantity) and P(D|H) interpreted as likelihood (an *a priori* given quantity). Bayesian inference comes in a familiar form to cognitive psychology: the Signal Detection Theory.

In the Signal Detection interpretation of the Bayesian formula Equation (1) above, the notions of hit, miss, false-alarm, and correct rejection outcomes are well defined in Table 1, and the ratios, HR (hit rate) and FAR (false-alarm rate) are defined with respect to the two hypotheses. Take the Mammography Problem as an example, HR (hit rate) is the percentage of disease cases caught by positive tests among all positive disease cases, whereas FAR (false-alarm rate) is the percentage of no-disease cases that came out with positive tests among all negative disease cases. In the language of Signal Detection Theory, this analysis is "stimulus-based", because the internal disease state ("disease" vs. "no disease") is mapped to the unknown "stimulus" state, and manifested "evidence" ("positive test result" or "negative test result") is mapped to the resulting "response." In the case of Signal Detection Theory, it turns out that we can perform "response-based" analysis (Zhang, Riehle, & Requin, 1997) and construct analogous hit rate and false-alarm rate conditioned on the evidence space. Specifically, defining the hit rate HR as the percentage of hits among all positive test results, and false-alarm rate FAR as the percentage of false alarms among all negative test results, HR is identified with P(H|D) and FAR as P(H|∼D):

$$HR = P(H|D) = \frac{a}{a+c}; \quad FAR = P(H|\sim D) = \frac{b}{b+d}$$

Then we derived an expression analogous to Equation (1) as:

$$\frac{a}{a+b} = \frac{\frac{a+c}{a+b+c+d} \times \frac{a}{a+c}}{\frac{a+c}{a+b+c+d} \times \frac{a}{a+c} + \frac{b+d}{a+b+c+d} \times \frac{b}{b+d}} \tag{2}$$

where

$$P(H|D) = \frac{a}{a+c}; \qquad P(H|\sim D) = \frac{b}{b+d};$$
$$P(D) = \frac{a+c}{a+b+c+d}; \quad P(\sim D) = \frac{b+d}{a+b+c+d}$$

In this case, an analogous formula results:

$$P(D|H) = \frac{P(D) \times P(H|D)}{P(D) \times P(H|D) + P(\sim D) \times P(H|\sim D)}.$$

Our observation above hints that Ω_{Hypo} and Ω_{Data} can be treated more or less on the same footing: both can be viewed as sample spaces that allow probability measures/masses to be defined on them. Moreover, we can view conditional probabilities as "transporters" that can transfer probability masses between those two sample spaces. The so-called "likelihood" function is then the "forward transporter" from Ω_{Hypo} to Ω_{Data}, whereas the "posterior probability" is the "backward transporter" from Ω_{Data} to Ω_{Hypo}; here "forward" and "backward" only have nominal meanings. It is the Bayesian theorem that requires the forward and backward transporter to behave consistently (Zhang, 2013).

This probability transfer viewpoint of Bayesian inference has its root in Optimal Transport Theory (Villani, 2008) which gives rise to the modern foundation of probabilistic sample spaces as metric measure spaces. Optimal transport problem originates from the so-called

Monge Problem (formulated by a French mathematician G. Monge in 1781) of moving mass from one place X to another place Y while minimizing the amount of work of transporting them. In the 1960s, the Soviet mathematician L. Kantorovich reconceptualized the Monge problem as being defined in a conjoined sample space $X \times Y$ constructed from the original space X and destination space Y, along with the joint probability measure defined on it subject to the marginalization constraints. In this article, we will not touch upon the optimality aspect of the theory, but merely borrow the analogy of probability transport.

Bayesian Reasoning as Probability Transport

The traditional interpretation of base rate neglect in Bayesian reasoning in humans is that base rate information represents the subject's prior belief, P(H), but priors need not equal base rates (Cosmides & Tooby, 1996). In fact, the prior reflects one's personal probabilistic assessment of H, given all that they know prior to learning D (Edwards, Lindman, & Savage, 1963); how people revise their beliefs or subjective probabilities in light of newly acquired evidence is not immediately available in human Bayesian reasoning task, so it may be unwise to conclude that people ignore base rate (Mandel, 2014). In the sequel, we treat prior probability merely as a probability measure defined on the hypothesis space that is specified while the Bayesian reasoning task instruction is given to the subjects; there is no temporal difference in acquiring its knowledge and the knowledge of likelihoods.

We now apply probability transport ideas to Bayesian reasoning task such as the classic Mammography Problem (Eddy, 1982). We assume that the task scenario invokes mental representations of two sets of uncertain events in an individual's mind, one involving disease/no disease (collectively forming the hypothesis space denoted as Ω_{Hypo}) and the other involving positive/negative test results (collectively forming the evidence or data space denoted Ω_{Data}). Both Ω_{Hypo} and Ω_{Data} are called *sample spaces*, because one can prescribe a measure of uncertainty, namely probability, on the sample values of the corresponding random variable (disease state or test result). Along with the construction of sample spaces in the mental model, we assume that an individual will also invoke the schemas that relate the two sample spaces, in particular, correlation between the set of events represented in one sample space and the set in the other. Causal inference may be invoked but is not necessary for associative correlation. Such association, presumably based on an individual's prior knowledge, is highly context-dependent. They establish some kind of bijective mapping between Ω_{Hypo} and Ω_{Data}. This is the basis for establishing support between elements of the two sample spaces, such that positive test result is linked with positive disease state. As each sample space comes with an uncertainty measure themselves (marginal probabilities), introducing specific evidence in Bayesian reasoning can be envisioned as transporting the support of such evidence (probability measure) from the evidence space to the hypothesis space; so the posterior probability is nothing but the original marginal probability measure as modified by the transport. The transport or transfer process itself may be, however, subject to perceptual filtering or memory retrieval effects. In particular, structural "relatedness" between Ω_{Hypo} and Ω_{Data} would play some role during the probability transport, by facilitating transfers in positively correlated scenario and inhibiting transfers in negatively correlated or uncorrelated scenarios.

More concretely, in the Mammography Problem familiar to Bayesian inference researchers, the available data, such as "positive test result" D, causes probability mass in the hypothesis space Ω_{Hypo} to redistribute due to transportation: the mass from its prior distributions $P(H)$ and $P(\sim H)$ to posterior distribution $P(H|D)$ and $P(\sim H|D)$. In relative odds format:

$$\frac{P(H)}{P(\sim H)} \rightarrow \frac{P(H|D)}{P(\sim H|D)}.$$

The transporter itself is the likelihood ratio: $P(D|H)/P(D|\sim H)$, that is, how the data D is explained by disease H or no-disease $\sim H$ hypothesis. However, instead of full Bayesian transport, which amounts to multiplying the prior odds and the likelihood ratio to generate posterior odds, we assume that human cognitive heuristics induce a facilitation/interference bias,

to be quantified by a positive factor η ($\eta > 0$). This leads the posterior odds to be

$$\frac{P(H|D)}{P(\sim H|D)} = \eta \cdot \frac{P(H)}{P(\sim H)} \cdot \frac{P(D|H)}{P(D|\sim H)}. \quad (3)$$

That is,

$$P(H|D) = \frac{\eta P(H)P(D|H)}{\eta P(H)P(D|H) + P(\sim H)P(D|\sim H)}. \quad (4)$$

In essence, the joint probability P(H, D) = P(H|D)P(H) has been expanded or shrunk by a factor of η, due to a cognitive facilitation/interference effect. Because η ranges from 0 to positive infinity, we transform it into a bounded variable $\alpha = \eta/(1 + \eta)$, which is called "coefficient of transport" ($0 < \alpha < 1$). A value of $\alpha = .5$ (corresponding to $\eta = 1$) indicates that the human reasoner neither facilitates nor interferes with the transport across sample spaces; this is the perfect Bayesian case.

We find it convenient to depict the process of Bayesian inference, and the associated facilitation/interference during probability transfer, using visual aids of a probability square to represent relative proportions as relative areas (cf. Bea, 1995). Figure 1A shows the two demarcations of the relevant probabilities (frequencies) according to different sample spaces: based on Ω_{Hypo} (left panel) and Ω_{Data} (right panel), reflecting the quantitative relationship expressed by Equation (1) and by Equation (2), respectively. This mental "shift" between these two frames: Ω_{Hypo} (frame of diseases) and Ω_{Data} (frame of test results) is accompanied by probability transfer across these two sample spaces. Normally, for a perfectly Bayesian reasoner, the area represented by P(H, D) is neither enlarged nor shrunk during this transfer. This leads to Equation (1) and Equation (2) both being valid expressions as the Bayes formula.

However, cognitive biases may affect this process of probability transportation. In the Mammography Problem (Figure 1B), we assume that participants have mentally represented a positive correlation between positive test result and positive disease state: They will expand the probability mass of those who received positive mammography and had breast cancer, compared with the probability mass of those who received positive mammography but had no breast cancer. Therefore, the joint probability P(H, D) will expand during probability transfer. Likewise, in another problem—the Lottery Problem, which has identical numerical information but different context, we assume that people only conceive weak relation, if at all based on their daily experience, between winning prize and ticket center at which the winning lottery ticket is sold (Shi & Zhang, 2009; see the materials of Experiment 1). Therefore, the sample space Ω_{Hypo} (win or not) is weakly related to Ω_{Data} (having bought from a certain ticket center or not), leading to no expansion of P(H, D) but rather coalignment of the two demarcation plots. The thick arrow represents the change of probability mass due to structural similarity for the reasoning task.

Empirical Testing of PTT

We will test the implication of our PTT model of Bayesian reasoning, namely, treating the reasoning as transporting probabilities between two sample spaces. To the extent that human reasoning is subject to cognitive heuristics and cognitive constraints, we would expect human Bayesian inference to deviate from normative Bayesian transportation in predictable ways. Two of such deviations will be investigated below.

The first potential deviation from normative Bayesian transportation is that there is no *a priori* reason to have the forward and backward transporters to be coupled in the way prescribed by the Bayes's theorem. Rather, transportation of probabilities between the two sample spaces would be affected by structural similarities, congruence, or otherwise relatedness. This strongly hints at contextual effect in human Bayesian reasoning. The second potential deviation is due to probability distortion intrinsic in human probabilistic calculus. It was named "'anchoring reference error' bias and 'intuitively adjust deviation'" (Shi & Zhang, 2009, 2011). We expect such distortion to manifest itself in the estimation of transported amount.

Our probability transfer viewpoint of Bayesian inference is consistent with prior work on role of samples space in uncertainty reasoning. Gavanski and Hui (1992) proposed that judgments of frequency or probability require formation of an appropriate sample space (i.e., consideration of the set of outcomes that could occur or that could be true). Hilbert (2012) went

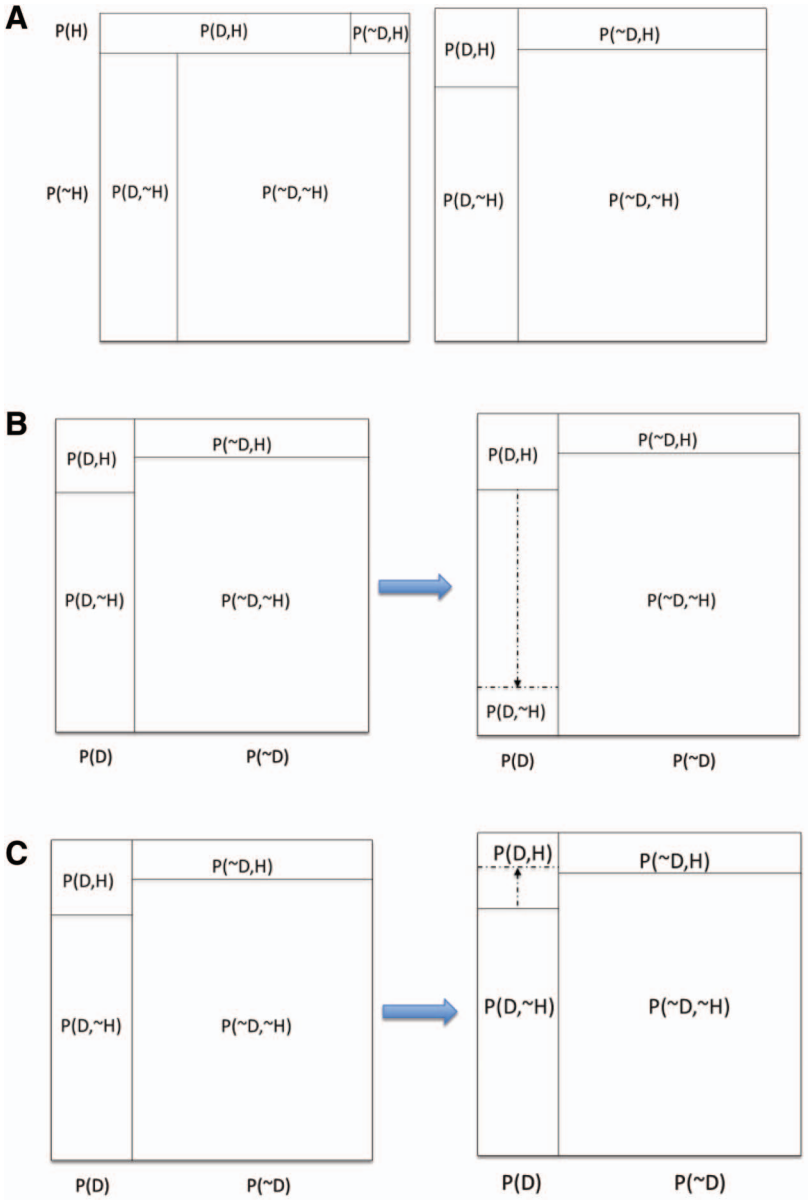

Figure 1. (A) Two demarcations of joint probabilities (frequencies) according to different sample spaces: based on Ω_{Hypo} (left panel) and Ω_{Data} (right panel). The left panel depicts the demarcation of probability of hypothesis space (H; horizontal line) first, followed by two vertical lines, whereas the right panel depicts the demarcation of probability of data space (D; vertical line) first, followed by two horizontal lines. All four areas in the left panel equal corresponding areas of the right panel (schematic drawing). In particular, area of P(D, H) is of same size in the two panels (for a ideal Bayesian reasoner). (B) Transfer between two sample spaces in the Mammography Problem. Dotted arrow represents the shift of relative probabilities due to high correlation (positive structural similarity) between D and H. (C) Transfer between two sample spaces in the Lottery Problem. Dotted arrow represents the shift of relative probabilities due to a low correlation (and hence almost independent marginal probabilities) between D and H. See the online article for the color version of this figure.

further to propose a conceptual synthesis of "the noisy memory channel" to explain what can go wrong when, for example, given the probability of each realization of H and of each transition probability P(D|H), one is able to calculate the probability of each value of D. Our theoretical novelty is to further propose that it is the transportation of probability measures across these sample spaces that underlies cognitive biases in Bayesian reasoning.

Our article investigates Bayesian reasoning by viewing such inference as transport of probability measures from evidence space D to hypothesis space H. We investigate how probability transport may be affected by distinctive structures of the two sample spaces as constructed from people's prior knowledge structure. We ask the question of how performance will be impacted by providing participants with different contexts ("cover stories") which invoke different correlations between the two sample spaces albeit with the same probability information. Our hypothesis is as follows:

1. If the structural correlation between the two sample spaces is strong and positive (e.g., in the Mammography Problem with disease/no disease and positive/negative test results), then when people are asked to evaluate the probability of disease given positive test results, they will enlarge the proportion of P(H, D), leading to overestimation of posterior probability. In this case the coefficient of transport α will be more than 0.5.
2. If the structural correlation between the two sample spaces is strong and negative (e.g., in some versions of the Office Problem, see below), then people will decrease the proportion of P(H, D), leading to underestimation of posterior probability. In this case, the coefficient of transport α will be less than 0.5.
3. If the structural correlation between the two sample spaces is weak (e.g., in the Lottery Problem with "winning lottery or not" and "having bought from a certain ticket center or not"), then people will be subject to this independence anchoring, that is, assuming the two sample spaces to be mutually independent. As a result, posterior probability should be the same as prior probability ("anchoring effect").

We conducted empirical studies to test the above predictions of our PTT account of Bayesian inference. Our experimentation consists of two stages: in Study 1, we use the standard scenario for Bayesian reasoning, the Mammography Problem, and compare it to a newly designed the "Lottery Problem" which has identical numbers but different correlation between the two sample spaces. In Study 2, to prevent possible contaminations of scenario (the cover story) by other factors, such as medical problem versus layperson problem (Siegrist & Keller, 2011), positive versus negative affective setting (Shi, Zhou, & Liu, 2012), sophistication in terminology, or even such criticisms that the Mammography Problem is not "adapted to the living environment of young people" (Binder et al., 2015), we designed some test materials (called the "Factory Problem") whereby the change between the scenarios is minimized while the degree of correlation between the hypothesis space and evidence space is still manipulated.

Study 1: "Mammography Problem" Versus "Lottery Problem"

In this study, we focused on the classic "Mammography Problem" (adapted from Eddy, 1982) and a newly designed "Lottery Problem." These two problems have different structural relatedness between the hypothesis space Ω_{Hypo} and the evidence space Ω_{Data}, more specifically, Ω_{Hypo} and Ω_{Data} are much more highly related in the Mammography Problem and much less related in the Lottery Problem.

In the Mammography Problem, the hypothesis space Ω_{Hypo} is the presence or absence of disease and the data space Ω_{Data} is positive or negative test outcome. In the Lottery Problem, the hypothesis space Ω_{Hypo} is "winning or not winning" lottery ticket and the data space Ω_{Data} is "place lottery ticket is bought." We postulate that participants will display more overestimation in the Mammography Problem, due to the high relatedness of the hypothesis space and data space, than in the Lottery Problem.

Method

Participants. Four hundred and forty-six Chinese undergraduates (287 females, 159 males) enrolled at a university in central China

served as participants in this experiment. Students participated in the experiments to fulfill a part of a class requirement and received a small token gift. Following Mellers and McGraw (1999), each person was randomly assigned to one of the two versions of the reasoning task (the Mammography Problem or the Lottery Problem).

Materials. We used two scenarios of a Bayesian reasoning problem, with the same numerical probability values. One was Gigerenzer and Hoffrage's (1995) probabilistic Mammography Problem (adapted from Eddy, 1982), and the other was a Lottery Problem designed for this experiment. To make the reasoning tasks comparable, we kept the same numerical information. The Lottery Problem is stated as follows:

> The chance of winning a lottery is 1%. 80% of those winning tickets were sold from a given ticket center; 9.6% of nonwinning tickets were sold from this center. Suppose there is a ticket sold from this center, what is the probability that this ticket will win the lottery? __%.

Procedure. Data were gathered in four classrooms when students were enrolled in an introductory psychology course. Each student was provided with a questionnaire booklet and was instructed to complete it individually with paper-and-pen and to explain how they made this evaluation. The booklet contained the Mammography Problem *or* the Lottery Problem, along with other filler tasks. The participants received all oral and written instructions in Chinese. Each task took approximately 10 min to complete.

Results

The correct Bayesian posterior probability is 7.8%, denoted 7.8 (without the percentage sign, same as below), and is taken to be the correct answer. Actual responses from subjects were collected and analyzed, giving the mean and standard deviation of posterior probability $M =$ 41.08, $SD =$ 35.3 (the Mammography Problem) and $M =$ 19.4, $SD =$ 29.9 (the Lottery Problem); there was significant difference in response to the two problems (Mann–Whitney, $U = 15103.00$, $p = .000 < .001$, $n = 446$.

To quantify the response pattern, we grouped the raw data into percentages. We treated answers from 7.0 to 8.0 (and in accordance with their explanations, see Gigerenzer & Hoffrage, 1995 for rationales of this Double Check procedure) to be correct. And we regarded all other responses as incorrect and categorized them as either overestimates or underestimates (Bramwell, West, & Salmon, 2006). Table 2 shows the percentages of different categories of responses.

In accordance with previous reports (Eddy, 1982; Gigerenzer & Hoffrage, 1995), most (more than 80%) responses were incorrect for the Mammography Problem: 65.5% of participants made overestimates (compared to 34.5% in the Lottery Problem, which is a significant difference: $\chi^2(1) = 41.63$, $p = .000 < .001$, $\varphi_c = 0.31$). On the other hand, 59.8% of participants gave underestimates in the Lottery Problem—compared to 27.1% in the Mammography Problem, which is a significant difference as well: $\chi^2(1) = 46.898$, $p = .000 < .001$, $\varphi_c = 0.32$.

The distributions of responses to the Mammography and the Lottery scenarios are depicted as histograms in Figure 2. Visual inspection reveals that more overestimates occurred in the Mammography Problem whereas more underestimates were produced in the Lottery Problem. Furthermore, there is a considerable amount of responses anchored at 1%, the prior probability.

We used the receiver operating characteristics (ROC) curve (see Figure 3), a technique adapted from Signal Detection Theory, to analyze the difference between the two response distributions. The ROC curve is constructed using cumulative frequencies of the responses for the two scenarios. The area under ROC curve is a measure of the difference of these two distributions—if there is no statistically difference, then the ROC curve would be the diagonal line, with an area of exactly ½ or 0.5. The area for the ROC curve associated with the ROC curve

Table 2
Percentage (Number) of Responses Which Were Correct, Overestimates, or Underestimates in Study 1

Response categories	The Mammography Problem	The Lottery Problem
Correct	7.4 (19)	5.7 (10)
Overestimates	65.5 (179)	34.5 (61)
Underestimates	27.1 (73)	59.8 (104)
Total	100 (271)	100 (175)

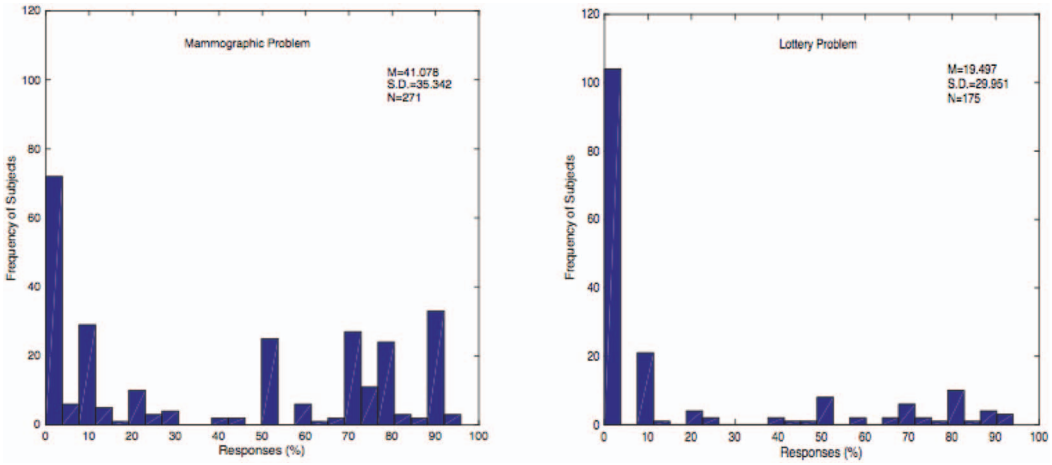

Figure 2. Percentages of participants giving responses in the Mammography Problem and in the Lottery Problem. See the online article for the color version of this figure.

generated from the Mammography Problem (X-axis) versus the Lottery Problem (Y-axis) comparison is 0.682.

Furthermore, we calculated the "coefficient of transport" α separately for each individual. The mean and media were respectively found to be 0.6601 and 0.9224 in the Mammography Problem, and 0.3886 and 0.1071 in the Lottery problem. Both were significantly different from the neutral value of 0.5 ($p < .01$). The relative frequency distributions of α value were shown in Figure 4.

To summarize, Study 1 shows that the "coefficient of transport" α is larger than 0.5 for the Mammography Problem, but this value is less than 0.5 in the Lottery Problem. That is, participants gave more overestimates in the Mammography Problem than in the Lottery Problem, and more underestimates in the Lottery Problem than in the Mammography Problem. Furthermore, the analysis of distribution of responses indicates that participants in the Lottery Problem are more likely to anchor at the prior probability, consistent with the view of independence of the hypothesis sample space (regarding winning ticket) and the data sample space (regarding ticket center). This pattern of over- and underestimates (with anchoring on prior) is consistent with the prediction of our PTT model based on structural relatedness of the hypothesis space and the data space in the two problems.

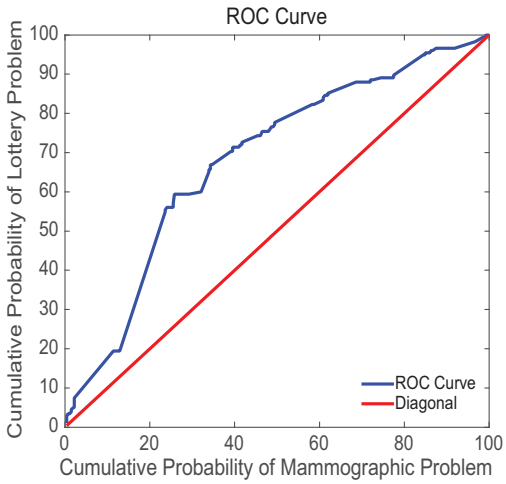

Figure 3. Receiver operating characteristic (ROC) curve for Study 1 (based on data in Figure 2). See the online article for the color version of this figure.

Study 2: Manipulating Contexts in a Single Problem

Study 1 demonstrates that the Mammography Problem and the Lottery Problem, though isomorphic to one another as a problem of estimating Bayesian posterior probability given prior probability and likelihood of evidence, gave rise to rather different response patterns. This was attributed to different degrees of correlation be-

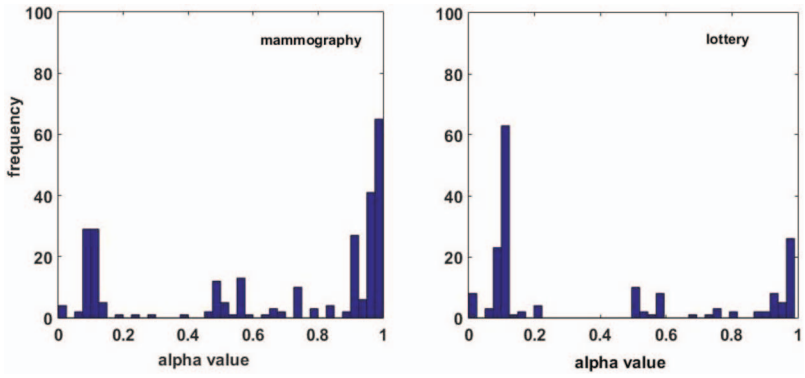

Figure 4. Distributions of alpha values in the Mammography Problem and in the Lottery Problem. See the online article for the color version of this figure.

tween the hypothesis space Ω_{Hypo} and the data space Ω_{Data}. However, the semantics of these 2 problems are quite different, so the different response pattern could be due to possible contaminations of scenario by other factors irrelevant to our conjecture, namely medical setting versus social–cognitive setting (Siegrist & Keller, 2011), positive versus negative affective frame (Shi et al., 2012), and so forth. To further test our Probability Transport model about how reasoning performance depends on prior assumptions about structural congruence between the two sample spaces Ω_{Data} and Ω_{Hypo}, we designed Study 2 in which there is a single scenario but with three alternative versions for Bayesian reasoning, all of which carrying the same probability numbers. Each version would, through invoking subjects' mental representations, different amount of structural similarities and hence correlations that affect probability transport. The inference problem is called the Factory Problem, which is about whether an employee of a factory is more likely to have received college-level (advanced) degree if their job is in the Research and Development (R&D) Department, the Staff Office, or the Mold Workshop, respectively. The three job responsibilities carry varying amount of correlation with educational training level—R&D and office are generally regarded as "white-collar" jobs, whereas mold workshop is deemed as "blue-collar" job. The prediction of our PTT is that the stronger correlation between Ω_{Data} and Ω_{Hypo}, the more likely overestimation of posterior probability will occur.

Method

Participants. One hundred and sixty-eight (168) students (90 Females, 78 males) enrolled at a university in central China served as participants in this experiment. Their ages ranged from 18 to 22 ($M = 20$, $SD = 1.23$). Other aspects of participant profile are like those in Study 1.

Materials. We designed three related stories of Bayesian reasoning, involving an employee (with or without an advanced degree) working at various departments of a factory. The wording reads as follows:

> In a factory, there are 1% employees who have advanced degrees (defined as college and above). According to statistics, 80% of employees with advanced degrees work in SUBSTITUTE, while 9.6% employees with lower education level work in SUBSTITUTE. Suppose Mr. Liming is an employee who works in SUBSTITUTE, what is the probability that he has an advanced degree? (Please use percentage to answer this question).

Here the scenario words "SUBSTITUTE" is substituted by "R&D department," "general office," "mold workshop" for the three versions of this factory scenario.

Procedure. Three different questionnaires, each with one kind of scenario words printed, were randomly distributed and completed by participants. They were allowed to use calculators in this pencil-and-paper test and were also asked to explain their reasoning after their answers. Each participate then answered a question about what percentage of advanced degree holders work in (their respective scenario word) workplace.

Results

As with Study 1, the correct Bayesian posterior is 7.8 (in percentage). The average of participants' answers were, respectively, 47.42 (in R&D version), 47.56 (office version), and 29.46 (mold workshop version). There was a significant difference among versions (Kruskal-Wallis, $H(2) = 8.934$, $p < .05$).

Like Study 1, we categorized the participants' estimates of posterior probability into correct and incorrect answers, with the latter further classified into overestimates and underestimates. As shown in Table 3, over 85% of participants made incorrect estimates; however, the pattern of under- versus overestimates differ across the three scenarios. As for overestimate, we found that there is marginally significant difference between office and mold workshop versions, $\chi^2(1) = 3.513$, $p = .061 < 0.10$, $\varphi_c = 0.18$, while the difference between R&D and office versions, between office and mold workshop versions were not statistically significant ($p > .05$).

Most responses were clustered around two specific values: 1 or 80. In the R&D version, most answers were larger than 80 ($n = 25$ or 45.3% of our sample, vs. in office task version $n = 20$ or 35.8%, vs. in mold workshop version $n = 8$ or 14.2%). In mold workshop version, most answers were less than 1 ($n = 12$ or 21.2% of our sample, vs. in R&D version $n = 9$ or 16.3%, vs. in office version $n = 7$ or 12.5%).

Figure 5 shows the relative frequency distributions of responses to the three versions, labeled as R&D and Office and Mold problem. More overestimates occurred in the R&D and in the Office versions, and more underestimates occurred in the Mold Workshop version.

We constructed ROC curves for each of the pairwise distributions (see Figure 6). The area under ROC curves is: 0.510 (R&D against Office scenario), 0.633 (R&D against Workshop scenario) and 0.649 (Office against Workshop scenario). An ROC area of 0.5 shows that there is no statistically difference between the two sets of distributions.

As in Study 1, we calculated individual's coefficient of transport α, with results shown in Table 4. Testing against the null hypothesis of $\alpha = .5$, there was a significant difference for the R&D scenario, $t(54) = 4.143$, $p = .000$, $d = 0.56$ and for the Office scenario, $t(55) = 4.917$, $p = .000$, $d = 0.66$, but there was no statistically significant difference in the Workshop scenario, $t(56) = 1.476$, $p = .146 > .05$. Using the median value of α as the index that distinguishes reasoners' performance in the three scenarios, they are 0.9223602 (in R&D scenario), 0.7243507 (in Office scenario), 0.5689655 (in Workshop scenario). The relative frequency distributions of α are shown in Figure 7.

Taken together, data from Table 3, Figures 5, 6, and 7 show that although R&D scenario and Office scenario do not generate statistically significant difference in the performance of Bayesian reasoning, both generate significantly more overestimates compared with the Mold Workshop Scenario.

Table 3
Percentage (Number) of Responses Which Were Correct, Overestimates, or Underestimates in Study 2

Response categories	R&D problem	Office problem	Mold workshop
Correct	14.5 (8)	8.9 (5)	14.2 (8)
Overestimates	63.8 (35)	71.4 (40)	54 (31)
Underestimates	21.7 (12)	19.7 (11)	31.8 (18)
Total	100 (55)	100 (56)	100 (57)

Conclusion and Discussions

In this article, we proposed the Probability Transfer Theory (PPT) of human Bayesian reasoning. PTT postulates that Bayesian inference involves transportation between two different sample spaces, namely, a hypothesis space and an evidence (or data) space. Individuals are constantly constructing and reconstructing probability measures (subjective beliefs of uncertainty) on these sample spaces based on available information as well as their mutual consistencies; this leads to the transportation of probabilities between sample spaces, where probability transfer (by "transporters") is subject to facilitation/interference due to structural similarities, congruence, or otherwise relatedness of the sample spaces involved. Representing the probability mass as the unit square (see Bea, 1995; Sturm & Eichler, 2014), we graphically depict the probability measures and their transporter (conditional probabilities) in Figure 1A. Our theory postulates dynamic interactions of putting probability measures on these spaces in

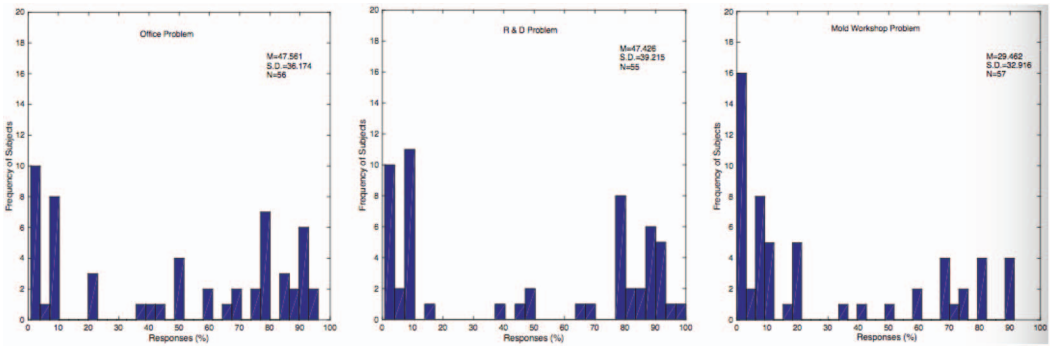

Figure 5. Percentages of participants giving responses in the three scenarios of the Factory Problem. See the online article for the color version of this figure.

context-dependent ways during any probabilistic inference task (Figure 1B and 1C).

We conducted two experiments to test our conjecture that the transporter (probability transfer) is influenced by the perceived correlation of the two sample spaces based on participants general knowledge of the task. Specifically, strong positive correlation is assumed to lead to expansion of the join probability measure P(H, D), whereas strong negative correlation will lead to shrinking that amount. The amount of expansion/shrinkage, reflecting cognitive facilitation/interference, depends on the degree to which the hypothesis space and evidence space are cor-

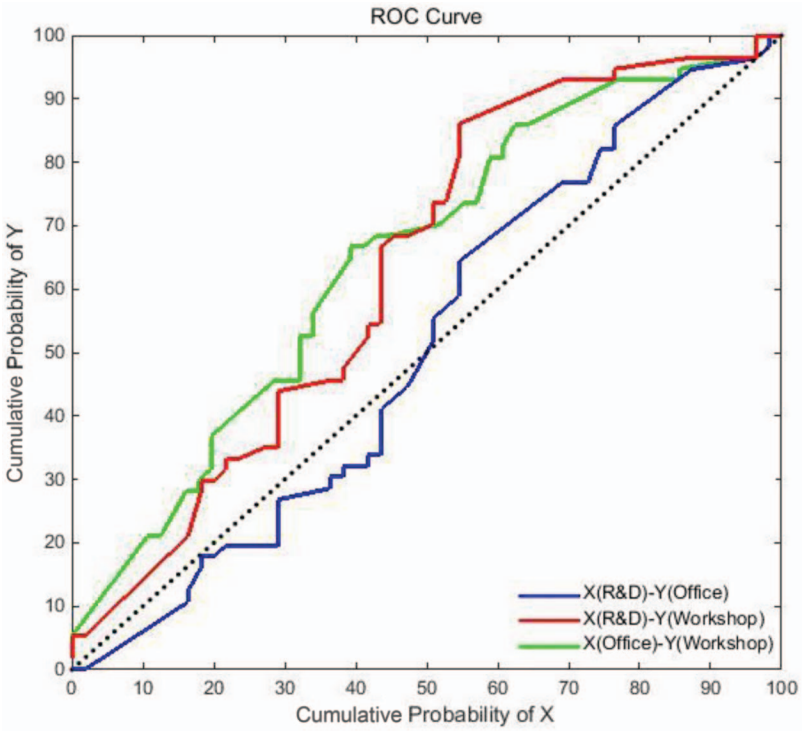

Figure 6. Receiver operating characteristic (ROC) curves for reasoning in the Factory Problem (based on data from Figure 5). See the online article for the color version of this figure.

Table 4
Mean and Standard Deviation of "Coefficients of Transport" α in Three Scenarios

Type of scenario	R&D	Office	Workshop
M ± *SD*	.696 ± .35	.724 ± .34	.575 ± .36

Note. R&D = research and development.

related; we quantify this amount by a coefficient of transport that we computed for each participant in our study. Compared with traditional data format for research in Bayesian inference, which for most part is only population mean/median and standard error, our refined ROC analysis borrowed from Signal Detection Theory attests to the robustness of our results. Our postulated relationship between positive/negative correlations of the two sample spaces and the transport coefficient is borne out. This pattern of results is consistent with the view (e.g., Mandel, 2014) that one cannot simply equate base rate with the subject's prior belief in the Mammography Problem (and other related Bayesian inference problems); base rate represents no more than a probability measure on one of the sample spaces and hence is partial information just as conditional probabilities (modeled as transporters in our PTT conceptualization).

Context is known to provide a vital influence on the outcome of Bayesian inference. In the study of Binder et al. (2015), the performance in almost every version of the economics problem (30% correct inferences, averaged across format of information and visualization) was much better in the Mammography Problem (correct rate 16%). They thought it might be the extreme base rate (1%) in the Mammography Problem which basically constitutes the cognitive illusion (in contrast, the result of the economics problem is no longer counterintuitive), and it might be that the context of the economics problem is more adapted to the living environment of young people (a strong dependency from the problem context was also found by Siegrist & Keller, 2011). They thought also that the more complicated terminology in the Mammography Problem could also account for the deviant effects in the different contexts. (e.g., Lesage, Navarrete, & DeNeys, 2013; Sirota, Juanchich, & Hagmayer, 2014). While acknowledging the importance of context, our PTT provides a framework for addressing the various loci where context can affect Bayesian reasoning, for instance, the independence/correlation of the sample spaces, different behaviors of the forward/backward transporter as modeling conditional probabilities, and so forth, Previously, Oaksford, Chater, and Larkin (2000) and Oaksford and Wakefield (2003) suggested that daily experience can influence people's estimate of conditional probabilities P(X|Y), both the X-part (conditioned) and the Y-part (conditioning) of the conditional rule. In an earlier study involving one of the current authors, Tang and Shi (2011) used the hepatitis and lottery as contexts and manipulated the levels of the base rate and the hit rate. Results indicate that the processing of hit rate information P(D|H) is strongly influenced by how/whether it is congruent with cognitive schemas of the participants as invoked by the cover

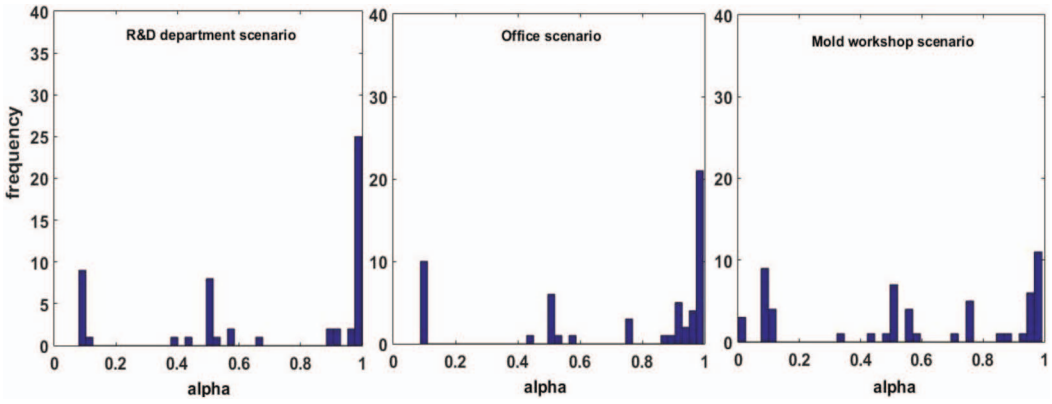

Figure 7. Distribution of alpha values in the R&D, Office and Mold workshop scenarios. See the online article for the color version of this figure.

story. Their quantitative implications are yet to be worked out.

In previous empirical studies, physicians' diagnostic inferences were thought to be based on prevalence as encoding base-rate information, and sensitivity and specificity of the test encoded as hit rate and false-alarm rate (Hoffrage & Gigerenzer, 1998). However, Steurer, Fischer, Bachmann, Koller, and ter Riet (2002) found that most physicians strongly overestimated the probability of disease given the positive test result—most doctors seem to be convinced by their experience about the correlation between decease and positive test result, and gave overestimation as a result. In fact, in Eddy's original study (1982), most of the physicians (95 out of 100) estimated the higher predictive value of the test. So it is possible the structural linkage between the hypothesis and evidence space may be particularly strong in experts compared with novices, leading to overestimation. We speculate that the facilitation/interference of probability transfer across sample spaces mostly invokes the System 1 ("associative system") in the Dual Process Theory (Barbey & Sloman, 2007; Kahneman & Frederick, 2005; Sloman, 1996). Future research will illuminate whether there is systematic difference in the way probability measures are being transferred across sample spaces for novices and for domain experts.

References

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences, 30,* 241–254. http://dx.doi.org/10.1017/S0140525X07001653

Bea, W. (1995). *Stochastisches Denken* [Stochastical Reasoning]. Frankfurt am Main, Germany: Peter Lang.

Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information – an empirical study on tree diagrams and 2 × 2 tables. *Frontiers in Psychology, 6,* 1186. http://dx.doi.org/10.3389/fpsyg.2015.01186

Bramwell, R., West, H., & Salmon, P. (2006). Health professionals' and service users' interpretation of screening test results: Experimental study. *British Medical Journal, 333,* 284–286.

Brase, G. L., Fiddick, L., & Harries, C. (2006). Participant recruitment methods and statistical reasoning performance. *The Quarterly Journal of Experimental Psychology, 59,* 965–976. http://dx.doi.org/10.1080/02724980543000132

Brase, G. L., & Hill, W. T. (2015). Good fences make for good neighbors but bad science: A review of what improves Bayesian reasoning and why. *Frontiers in Psychology, 6,* 340. http://dx.doi.org/10.3389/fpsyg.2015.00340

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1–73. http://dx.doi.org/10.1016/0010-0277(95)00664-8

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). NewYork, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511809477.019

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York, NY: Wiley.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242. http://dx.doi.org/10.1037/h0044139

Ellis, K. M., Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (2014). Do people understand their home HIV test results? Risk literacy and information search. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58,* 1323–1327. http://dx.doi.org/10.1177/1541931214581276

Fiedler, K., Brinkmann, B., Betsch, T., & Wild, B. (2000). A sampling approach to biases in conditional probability judgments: Beyond base rate neglect and statistical format. *Journal of Experimental Psychology: General, 129,* 399–418. http://dx.doi.org/10.1037/0096-3445.129.3.399

Gavanski, I., & Hui, C. (1992). Natural sample spaces and uncertain belief. *Journal of Personality and Social Psychology, 63,* 766–780. http://dx.doi.org/10.1037/0022-3514.63.5.766

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704. http://dx.doi.org/10.1037/0033-295X.102.4.684

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. *Cognition, 78,* 247–276. http://dx.doi.org/10.1016/S0010-0277(00)00133-5

Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin, 138,* 211–237. http://dx.doi.org/10.1037/a0025940

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Ac-*

ademic Medicine, 73, 538–540. http://dx.doi.org/10.1097/00001888-199805000-00024

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Medicine. Communicating statistical information. *Science, 290,* 2261–2262. http://dx.doi.org/10.1126/science.290.5500.2261

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50,* 109–135. http://dx.doi.org/10.1146/annurev.psych.50.1.109

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. J. Holyoak & R. G. Morris (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology, 3,* 430–454. http://dx.doi.org/10.1016/0010-0285(72)90016-3

Kleiter, G. D. (1994). Natural sampling: Rationality without base rates. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 375–388). New York, NY: Springer. http://dx.doi.org/10.1007/978-1-4612-4308-3_27

Lesage, E., Navarrete, G., & DeNeys, W. (2013). Evolutionary modules and Bayesian facilitation: The role of general cognitive resources. *Thinking & Reasoning, 19,* 27–53. http://dx.doi.org/10.1080/13546783.2012.713177

Mandel, D. R. (2014). The psychology of Bayesian reasoning. *Frontiers in Psychology, 5,* 1144. http://dx.doi.org/10.3389/fpsyg.2014.01144

Mellers, B. A., & McGraw, A. P. (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage (1995). *Psychological Review, 106,* 417–424. http://dx.doi.org/10.1037/0033-295X.106.2.417

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 883–899. http://dx.doi.org/10.1037/0278-7393.26.4.883

Oaksford, M., & Wakefield, M. (2003). Data selection and natural sampling: Probabilities do matter. *Memory & Cognition, 31,* 143–154. http://dx.doi.org/10.3758/BF03196089

Phillips, L. D., & Edwards, W. (1966). Conservatism in a simple probability inference task. *Journal of Experimental Psychology, 72,* 346–354. http://dx.doi.org/10.1037/h0023653

Shi, Z. F., & Zhang, Q. L. (2009). The effect of anchoring reference error biases on Bayesian reasoning [in Chinese]. *Psychological Science, 32,* 446–448.

Shi, Z. F., & Zhang, Q. L. (2011). The effect of intuitively adjust deviation biases on Bayesian reasoning [in Chinese]. *Psychological Science, 34,* 970–973.

Shi, Z. F., Zhou, Y. X., & Liu, M. (2012). The influence of different task context and emotion states on Bayesian reasoning [in Chinese]. *Psychological Science, 35,* 988–992.

Siegrist, M., & Keller, C. (2011). Natural frequencies and Bayesian reasoning: The impact of formal education and problem context. *Journal of Risk Research, 14,* 1039–1055. http://dx.doi.org/10.1080/13669877.2011.571786

Sirota, M., Juanchich, M., & Hagmayer, Y. (2014). Ecological rationality or nested sets? Individual differences in cognitive processing predict Bayesian reasoning. *Psychonomic Bulletin & Review, 21,* 198–204. http://dx.doi.org/10.3758/s13423-013-0464-6

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22. http://dx.doi.org/10.1037/0033-2909.119.1.3

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes, 91,* 296–309. http://dx.doi.org/10.1016/S0749-5978(03)00021-9

Steurer, J., Fischer, J. E., Bachmann, L. M., Koller, M., & ter Riet, G. (2002). Communicating accuracy of tests to general practitioners: A controlled study. *British Medical Journal, 324,* 824–826. http://dx.doi.org/10.1136/bmj.324.7341.824

Sturm, A., & Eichler, A. (2014, July). *Students' beliefs about the benefit of statistical knowledge when perceiving information through daily media*. Paper presented at the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ.

Tang, Y. H., & Shi, Z. F. (2011). The inhibition from hit rate reference of Bayesian reasoning [in Chinese]. *Psychological Science, 1,* 220–224.

Villani, C. (2008). *Optimal transport: old and new* (Vol. 338). Berlin, Germany: Springer Science & Business Media.

Yamagishi, K. (2003). Facilitating normative judgments of conditional probability: Frequency or nested sets? *Experimental Psychology, 50,* 97–106. http://dx.doi.org/10.1026//1618-3169.50.2.97

Zhang, J. (July, 2013). *Bayes theorem, mutual information, and Shannon source/channel coding: An information geometric perspective*. Oral presentation at the 46th Annual Meeting of the Society for Mathematical Psychology, Potsdam, Germany.

Zhang, J., Riehle, A., & Requin, J. (1997). Analyzing neuronal processing locus in stimulus-response association tasks. *Journal of Mathematical Psychology, 41,* 219–236. http://dx.doi.org/10.1006/jmps.1997.1168

Received January 19, 2016
Revision received January 29, 2019
Accepted March 13, 2019 ■