

Cite as:

Mangiarulo, M., Pighin, S., Polonio, L., & Tentori, K. (2020). The effect of evidential impact on perceptual probabilistic judgments. *Cognitive Science*, in press.

The effect of evidential impact on perceptual probabilistic judgments

Marta Mangiarulo¹, Stefania Pighin¹, Luca Polonio², Katya Tentori¹

¹ Center for Mind/Brain Sciences, University of Trento

² IMT school for Advanced Studies Lucca

Corresponding author:

Katya Tentori
Center for Mind/Brain Sciences (CIMEC)
University of Trento
Corso Bettini, n. 31 38068 Rovereto (TN), Italy
Email: katya.tentori@unitn.it

Keywords: evidential impact; Bayesian confirmation; probabilistic reasoning; visual features.

Abstract

In a series of three behavioral experiments, we found a systematic distortion of probability judgments concerning elementary visual stimuli. Participants were briefly shown a set of figures that had two features (e.g., a geometric shape and a color) with two possible values each (e.g., triangle or circle and black or white). A figure was then drawn, participants were informed about the value of one of its features (e.g., that the figure was a “circle”) and had to predict the value of the other feature (e.g., whether the figure was “black” or “white”). We repeated this procedure for various sets of figures and, by varying the statistical association between features in the sets, we manipulated the probability of a feature given the evidence of another (e.g., the posterior probability of hypothesis “black” given the evidence “circle”) as well as the support provided by a feature to another (e.g., the impact, or confirmation, of evidence “circle” on the hypothesis “black”). Results indicated that participants’ judgments were deeply affected by impact, although they only should have depended on the probability distributions over the features, and that the dissociation between evidential impact and posterior probability increased the number of errors. The implications of these findings for lower- and higher-level cognitive models are discussed.

Introduction

Within the Bayesian framework for modelling inductive reasoning, the relation between evidence (e) and hypothesis (h) is captured by two distinct notions: the *posterior probability* of the hypothesis in light of the evidence, $Pr(h|e)$, and the *impact* (or degree of *confirmation*) of the evidence on the hypothesis, $Imp(h,e)$. Consider, for example, a set of figures comprising 12 white circles, 12 black circles, 20 white triangles, and 4 black triangles. Imagine that a figure is randomly and covertly drawn from this set. Its shape is then revealed: the figure is a circle (e). You might be interested in updating the probability that the drawn figure is white (h) or black (not- h) in light of this new information, and/or in considering how knowing that the figure is a circle affects the credibility of the two alternative hypotheses at issue. While the former is a probability judgment, the latter is an impact judgment.

The *qualitative* notion of impact is defined as follows:

$$(1) \text{ } Imp(h, e) \begin{cases} \geq 0 \text{ iff } Pr(h|e) \geq Pr(h); \\ = 0 \text{ iff } Pr(h|e) = Pr(h); \\ \leq 0 \text{ iff } Pr(h|e) \leq Pr(h). \end{cases}$$

This means that evidence e has a *positive* [*negative*] impact on hypothesis h iff the posterior probability of the hypothesis h in light of evidence e is *higher* [*lower*] than the prior probability of h . In case posterior and prior probability coincide, e is *irrelevant* with respect to h . In the above example, the posterior probability of the two hypotheses given the evidence is equal, $Pr(\text{white}/\text{circle}) = Pr(\text{black}/\text{circle}) = .5$, however, the evidence that the figure is a circle is not neutral with respect to these hypotheses: it strengthens the probability that the figure is black while it weakens the probability that the figure is white (indeed $Pr(\text{black}) = .33 < .5 < .67 = Pr(\text{white})$). One might also be interested in quantifying impact, that is, to assess the amount of confirmation/disconfirmation that e brings to h . This has been addressed in a variety of measures (for a review see Fitelson, 1999; Crupi, Tentori, & Gonzalez, 2007). In our paper, we will make use of one prominent model that has been demonstrated to capture participants' naïve

impact judgments (Tentori, Crupi, Bonini, & Osherson, 2007) and corresponds to a monotone transformation of the *likelihood ratio* in the bounded interval $[-1,+1]$:

$$(2) \text{ Impl}(h, e) = \frac{\text{Pr}(e | h) - \text{Pr}(e | \neg h)}{\text{Pr}(e | h) + \text{Pr}(e | \neg h)} \quad (\text{Kemeny \& Oppenheim, 1952; Good, 1984}).$$

In most real-life circumstances, posterior probability and impact are positively correlated: when evidence strongly supports [opposes] a hypothesis, then the probability of the latter in light of the former is usually rather high [low]. However, posterior probability and evidential impact can be dissociated, as is often the case when the prior is extreme. For example, the probability that a man in his fifties has prostate cancer remains low even in light of evidence in its favor such as a positive PSA screening test result. Similarly, the probability that it will snow in Helsinki during the next winter remains high in spite of evidence pointing in the opposite direction, such as news that the coming winter in northern Europe will be particularly mild.

As clarified by the above examples, there are two key differences between evidential impact and posterior probability that are worth emphasizing here. First, while evidential impact is by definition a relative notion (it expresses the relevance of e with regards to h), posterior probability is an absolute value (it represents the “arrival point” of the belief in h once e is known). Second, the assessment of these two quantities is generally required in different reasoning tasks. The evaluation of impact has an important, albeit often implicit, role in all the situations where the value of information or the soundness of arguments has to be assessed, while the evaluation of posterior probability is fully acknowledged as a crucial component of various cognitive processes, including decision-making under uncertainty and forecasting. Given the importance, also in economic terms, of the latter activities, it is not surprising that historically the notion of posterior probability has received more attention than that of impact.

Over the last few decades, many empirical findings have questioned the accuracy of laypeople’s and experts’ intuitive probability judgments by showing their systematic departure

from normative benchmarks: the axioms of probability theory and their consequences, such as Bayes' theorem. Documented errors include, among others: conservatism (Edwards, 1968), base-rate neglect (Kahneman & Tversky, 1973; Tversky & Kahneman, 1982), order effects (Hogarth & Einhorn, 1992), conjunction (Tversky & Kahneman, 1983) and disjunction (Bar-Hillel & Neter, 1993) fallacies. Studies that considered various Bayesian models of inductive reasoning but also learning, categorization, language processing, and even lower-level cognitive processes, such as sensorimotor control and perception, provided more encouraging results by showing that cognitive behavior is often aligned with the “real” statistics of the environment (e.g., Chater, Oaksford, Hahn, & Heit, 2010; Kemp & Tenenbaum, 2009; Körding & Wolpert, 2006; Shen & Ma, 2016; Tenenbaum, Griffiths, & Kemp, 2006; Yuille & Kersten, 2006). However, most researchers agree now that such models should be intended as tools for testing psychological explanations of behavior rather than proper descriptions of how high and low-level probabilistic inferences work (see, for example, Griffiths, Chater, Norris, & Pouget, 2012; Marcus & Davies, 2013; Tauber, Navarro, Perfors, & Steyvers, 2017).

The empirical study of impact judgments is more recent, but data have converged in suggesting a good alignment with normative models. Indeed, participants' judgments of evidential impact consistently proved to be accurate both when referring to artificial material (e.g., opaque urns with balls of different colors, Tentori, Crupi, Bonini, & Osherson, 2007) and a variety of real-world predicates (e.g., “to be a male/female” and “to like ice-figure skating”, “to support a football team”, etc., Tentori, Chater, & Crupi, 2016). Sound judgments have also been found in tasks in which participants had to quantify the impact of uncertain evidence (Mastropasqua, Crupi, & Tentori, 2010) or the value of evidence with regards to competing hypotheses (Crupi, Tentori, & Lombardi, 2009; Rusconi, Marelli, D'Addario, Russo, & Cherubini, 2014). These results are consistent with those from the study of category-based induction, according to which not only adults, but also young children from 5 years old, evaluate

argument strength by following popular principles of evidential impact, such as the similarity between premises and conclusions or the diversity of premises (Heit & Hahn, 2001; Lo, Sides, Rozelle, & Osherson, 2002; López, Gelman, Gutheil, & Smith, 1992; Osherson, Smith, Wilkie, López, & Shafir, 1990; Zhong, Lee, Huang, & Mo, 2014).

Taken together, the above-mentioned empirical results suggest that human inductive ability might rely more on estimating evidential impact than posterior probability. This issue was directly explored by comparing the two kinds of assessments on the very same experimental materials (Tentori et al., 2016). Results indicated that impact judgments were more accurate (i.e., closer to objective values) and consistent across time (in a test-retest interval of one week) than corresponding posterior probability judgments.

The intuitive assessment of impact or, in other terms, the perceived relevance of information, has been proved to be a key determinant of classical probability reasoning errors, such as the conjunction fallacy (Bhatia, 2017; Tentori & Crupi, 2012; Tentori, Crupi, & Russo, 2013) and the base-rate neglect (Kahneman & Tversky, 1973; Bar-Hillel, 1980; Turpin et al., 2020). Similarly, the spontaneous and implicit appreciation of evidential impact has been shown to play a role, often under other names, in a variety of linguistic and psycholinguistic phenomena that have inherent probabilistic components, including representation of word meanings (Bullinaria & Levy, 2007), semantic priming (Nadalini, Marelli, Bottini, & Crepaldi, 2018), and estimation of the word co-occurrence likelihood (Paperno, Marelli, Tentori, & Baroni, 2014).

The aim of the present study is to examine whether impact relations can affect probability judgments even when the link between evidence and hypotheses is completely new and low-level, as in the case of the arbitrary association between visual features in briefly presented figures. This is a much stricter test than those already carried out since it looks for a systematic distortion in probability judgments that is not mediated by any consolidated semantic

knowledge or linguistic competence. Moreover, the use of abstract perceptual material provides a powerful way to fully disentangle impact and posterior. As these two quantities are often positively correlated in natural environments, one typically provides precious information on the other. As a consequence, the influence of impact on probability judgments referring to familiar stimuli might somehow be epistemically justified. The employment of novel artificial relations between two discrete perceptual features offers, instead, the possibility to document a net effect of impact over probability judgments that goes beyond these kinds of pragmatic concerns. Additionally, the results of our study may have implications relevant to perceptual research by contributing to a better understanding of how people create and manipulate their experience of physical events. Indeed, knowing whether and to what extent implicit detection of impact relations affects probability expectations regarding elementary visual stimuli would help inform us about how contingencies between basic visual features are represented, and might contribute to the current debate on probabilistic models of perception by specifying new elements that influence the accuracy of percepts (for a specific discussion on the importance of clarifying the accuracy-conditions in Bayesian perceptual psychology, see for example Rescorla, 2015).

General method

Three behavioral computer-based experiments were conducted using the same methodology. All experimental stimuli were controlled using MATLAB and the Psychophysics toolbox.

Participants

Participants were 104 students of the University of Trento (40, 32, and 32 for Experiments 1, 2, and 3 respectively). Demographic characteristics of the sample are described in the left part of Table 1. The study was approved by the Human Research Ethics committee of the University of Trento, and informed consent was obtained for all participants.

Stimuli and procedure

In all experiments, we employed 40 different sets of figures made of two features (e.g., a geometric shape and a color in Exp. 1) that could take two possible values each (e.g., triangle vs. circle for the geometric shape and black vs. white for the color). The kinds of figures, their overall number per set, and their presentation time were kept constant within an experiment but varied across experiments (details are provided in the right part of Table 1).

Table 1

Overview of the characteristics of the participants and the sets of figures used in the three experiments

Exp.	Participants		N per set	Presentation time	Figures	
	N (Female %)	Mean age (SD)			Feature 1	Feature 2
1	40 (75%)	23.9 (2.8)	46 or 48	3s	Color Black (rgb 0 0 0) White (rgb 255 255 255)	Shape Circular Triangular
2	32 (69%)	20.4 (2.2)	23 or 24	4s	Pattern Striped (thickness: 0.07° VA) Dotted (diameter: 0.17° VA)	Shape Circular Triangular
3	32 (84%)	22.4 (3)	23 or 24	4s	Curvature Wavy (thickness: 0.2° VA) Straight (thickness: 0.2° VA)	Orientation Horizontal Vertical

Participants were seated at a 60cm-distance from the computer screen (resolution: 1920 x 1080). Triangles and circles were inscribed in a (transparent) square of 2.6° visual angle (VA) per side, and they were randomly displayed at a minimum distance of 0.4° VA from each other. The background of the screen was mid gray in Experiment 1 (rgb 128 128 128) and light gray in Experiments 2 and 3 (rgb 200 200 200).

Trials proceeded as follows (see also Figure 1): A set of figures was displayed on the screen for 3s (Experiment 1) or 4s (Experiments 2 and 3); a figure was then drawn; participants were informed about one of its features (which served as evidence, for example, “triangle”) and had to predict the value of the other feature by choosing between the two possible alternatives (which counted as hypotheses, for example, “black” vs. “white”). The position of the figures in the sets was randomized, as was the presentation order of the trials, with a 30s break after every 20 trials.

The experimental trials were preceded by four training trials. As a test of consistency,

participants were presented with the same sets of figures and the same evidence (i.e., they were asked the same judgment) twice, for a total of 80 trials (= 40 stimuli X 2 presentations per stimulus).

Stimuli were generated in order to fully balance posteriors and impacts. More specifically, by manipulating the statistical association between the values of the features in the sets as well as the features that served as evidence and hypothesis, we obtained 20 pairs of arguments with the same posterior (e.g., $Pr(\text{black}|\text{triangle}) = .6$) but opposite impacts, i.e., a confirmation and a disconfirmation which were equal in absolute strength according to equation (2) (e.g., $Imp(\text{black}, \text{triangle}) = +.4$ and $Imp(\text{black}, \text{triangle}) = -.4$). Moreover, we imposed that each value of the two features (e.g., triangle) served as evidence one fourth of the time, in half of which it confirmed and in the other half disconfirmed to the same extent each level of the other feature, while posterior probability was held constant.

Participants were tested individually. They were instructed to be as fast and accurate as possible and had 30s to submit their answer. No immediate feedback on the accuracy of responses was provided; at the end of the task, participants earned €0.15 for each correct prediction.

Participants were not asked to report their strategies. Note, however, that counting was almost impossible, as it would have meant to independently update four different numbers, one for each of the relevant conjunctions. Indeed, when exploring the set of figures participants did not know yet which one of the two features or which value of this specific feature they would have received as evidence (each of the four possibilities had equal prior probability, because as explained above, each value of the two features served as evidence exactly one fourth of the times).

Specific verbal instructions, all experimental stimuli, and comprehensive data sets have been deposited in the Open Science Framework (doi: 10.17605/OSF.IO/QNEMY).

Data analysis

Data analysis was conducted using SPSS statistical analysis software (version 23). The same analytical strategy was used for all the three experiments. Since each participant contributed to many data points, we employed generalized mixed models that take this non-independence into account by adding random effects for participants.

First, to investigate whether participants' probability judgments were affected by evidential impact, we developed a binomial generalized linear mixed model on participants' choice for the hypothesis that was confirmed versus disconfirmed by the evidence, including random intercept for participants (we will refer to this analysis as GLMM₁). Then, to see whether participants' choices for the confirmed versus disconfirmed hypothesis were affected by the specific feature that served as evidence (e.g., shape or color in Exp. 1), we included, in the previous model, this dichotomic variable (we will label it type of evidence) both as a fixed effect and as random slope for participants (we will refer to this analysis as GLMM₂). Similarly, to test whether the type of evidence affected the time required to make a choice we ran a generalized linear mixed model on response times (RTs) using a logarithmic transformation of the values in order to approximate a normal distribution (Newell & Rosenbloom, 1981). We included the type of evidence in the model both as a fixed effect and random slope for participants, along with random intercepts for participants (GLMM₃).

Only for trials in which posterior probability differed from .5 (that is, trials in which there was a normatively correct response), we first checked whether the most and least probable hypotheses were equally chosen by developing a binomial generalized linear mixed model on participants' errors, including random intercept for participants (we will refer to this analysis as GLMM₄). Then we compared the frequency of errors (i.e., the choices for the less probable hypothesis) when the more probable hypothesis was confirmed versus disconfirmed by the

evidence (i.e., when $Pr(h|e) > .5$ and $Imp(h, e) > 0$ vs. $Pr(h|e) > .5$ and $Imp(h, e) < 0$). To this aim, we performed a GLMM analysis (GLMM₅) on participants' errors by including the agreement between posterior probability and impact (we will label this variable *Pr-Imp* agreement) both as a fixed effect and as random slope for participants, along with random intercept for participants. Finally, a GLMM analysis (GLMM₆) was conducted to evaluate if the agreement between probability and impact was a significant predictor of the decision time. This last model investigated participants' RTs as dependent variable and included *Pr-Imp* agreement as fixed factor and random slope for participants, along with random intercept for participants (a logarithmic transformation of the RT values was applied).

The same six analyses described above were also run for consistent trials only, that is for trials in which participants gave the same answer to the two occurrences of the same set of figures and evidence. Since the results of the analyses on consistent trials duplicate in all essential details the results obtained with all trials, we will report only the latter.

The proportion of choice for the confirmed hypothesis and the proportion of errors (for all trials and for consistent trials only) in the three experiments are presented in Tables 2 and 3, respectively. Main statistics of the GLMM analyses are reported in Table 4. Figure 2, 3, and 4 displays, for Experiment 1, 2, and 3, respectively, the proportions, at participant-level, of choices for the confirmed hypothesis and of errors when posterior probability and impact agree versus disagree.

Table 2

Choices of the confirmed hypothesis in the three experiments

Experiment 1				Experiment 2				Experiment 3			
Evidence	All trials	Consistent trials		Evidence	All Trials	Consistent trials		Evidence	All trials	Consistent Trials	
	Conf	Conf	<i>n</i>		Conf	Conf	<i>n</i>		Conf	Conf	<i>n</i>
Color	.70	.78	.65	Pattern	.68	.79	.61	Orientation	.64	.72	.62
Shape	.45	.41	.56	Shape	.49	.49	.62	Curvature	.61	.66	.61
Overall	.57	.61	.61	Overall	.59	.64	.61	Overall	.63	.70	.62

“Conf” stands for the proportion of choices favoring the confirmed hypothesis in all trials (column “All trials”) or in consistent trials only (column “Consistent trials”). These proportions have been computed by averaging across participants the proportion of choices for the confirmed hypothesis provided by each participant in trials in which the two features served as evidence. “*n*” stands for the average proportion of consistent trials out of the total amounts of trials.

Table 3

Proportion of errors in the three experiments

Experiment 1				Experiment 2				Experiment 3			
<i>Pr-Imp</i> Agreement	All trials	Consistent trials		<i>Pr-Imp</i> Agreement	All Trials	Consistent trials		<i>Pr-Imp</i> Agreement	All trials	Consistent Trials	
	Err	Err	<i>n</i>		Err	Err	<i>n</i>		Err	Err	<i>n</i>
Disagreement	.48	.47	.62	Disagreement	.49	.46	.60	Disagreement	.53	.54	.61
Agreement	.33	.25	.63	Agreement	.30	.19	.65	Agreement	.27	.17	.66
Overall	.41	.35	.62	Overall	.39	.32	.62	Overall	.40	.34	.64

“Err” stands for the proportion of errors in trials in which posterior probability and impact disagreed (row “Disagreement”) or agreed (row “Agreement”), when all trials (column “All trials”) or only consistent trials (column “Consistent trials”) were considered. These proportions have been computed by averaging across participants the proportion of errors made by each participant in which posterior probability and impact disagreed or agreed. “*n*” stands for the average proportion of consistent trials out of the total amounts of trials in which posterior probability differed from .5 (that is, trials in which there was a normatively correct response).

Experiment 1

Experiment 1 represented a first exploration of whether and how evidential impact might affect probability judgments when simple perceptual features like color (black vs. white) and shape (triangle vs. circle) were at issue.

Results

The first analysis (GLMM₁) revealed that participants were significantly more likely to choose the confirmed hypothesis (mean = .57; *SD* = .14; 95% CI = .53-.62) than the (on average equally probable) disconfirmed hypothesis, $\text{Exp}(B) = 1.37$; CI = 1.14-1.64. The proportion of

choices favoring the confirmed hypothesis (across all types of evidence) is displayed in the upper panel of Figure 2. The GLMM₂ showed that the type of evidence was a significant predictor of participants' choices for the confirmed hypothesis, $F(1, 3,198) = 55.7, p < .001$. Specifically, participants more frequently selected the confirmed hypothesis when the evidence was color (mean = .70; $SD = .17$) rather than shape (mean = .45; $SD = .17$). The GLMM₃ analysis revealed a significant main effect of the type of evidence on RTs, $F(1, 3,198) = 17.69, p < .001$. Specifically, RTs were longer when the type of evidence was color (mean = 2,208ms; $SD = 645$ ms) rather than shape (mean = 2,056ms; $SD = 737$ ms).

GLMM₄ revealed that, overall, the less probable hypotheses were less likely to be chosen (mean = .41; $SD = .01$; 95% CI = .39-.43) than the more probable ones, $\text{Exp}(B) = .686$; CI = .63-.75. The agreement between posterior probability and impact significantly predicted participants' errors, $F(1, 2,558) = 20.08, p < .001$. Indeed, the number of errors was greater when the more probable hypothesis was disconfirmed (mean = .48; $SD = .15$) rather than confirmed (mean = .33; $SD = .15$); the proportion of errors in the two conditions for each participant is reported in the lower panel of Figure 2. As revealed by GLMM₆, however, the time required to make a choice was not significantly affected by the agreement between posterior probability and impact, $F(1, 2,558) = .336, p = .562$ (mean = 2,153ms, $SD = 726$ ms and mean = 2,121ms, $SD = 639$ ms, for disagreement and agreement between posterior probability and impact, respectively).

To sum up, in Experiment 1, there was an overall effect of impact over probability judgments: participants chose more often the confirmed hypotheses over the (on average equally probable) disconfirmed ones and made more errors (i.e., they chose the less probable hypothesis) when posterior probability and impact were dissociated (i.e., when the more probable hypothesis was disconfirmed by the evidence). There was also an effect of the feature that served as evidence: participants chose more frequently the confirmed hypothesis when they

were provided the color of the figure and had to predict its shape rather than when it was the other way around. Such an asymmetry could originate from a different salience of the two features used (at least as regards the levels through which we have operationalized them: white vs. black and triangle vs. circle). In line with the empirical findings that indicate a privileged role of color in visual search (Wolfe & Horowitz, 2004, 2017) as well as in modulating attentional capture and unconscious priming (Adamo, Wozny, Pratt, & Ferber, 2010; Breitmeyer, Ogmen, & Chen 2004), color might have been a more compelling factor than shape in determining participants' responses. More specifically, it is possible that, when the color of the figure was provided as evidence, participants exploited it for updating their prediction concerning the shape of the figure, whereas, when the evidence was the shape of the figure, judgments were driven mainly by the prior probabilities of two alternative colors, without these being updated on information regarding shape. The RTs analysis is coherent with this interpretation: participants' judgments were faster when the color was used as hypothesis rather than evidence, which is compatible with the former judgments being driven only by priors, while the latter being the results of a probabilistic updating.

Experiment 2

In order to balance the salience of the two perceptual features used in the stimuli of Experiment 1, in Experiment 2 we replaced color with pattern (lines vs. dots, see Figure 1). Moreover, to simplify the task and reduce possible noise, we halved the number of figures in the sets and incremented the presentation time of each trial (see Table 1).

Results

As in Experiment 1, the first analysis (GLMM₁) showed that participants were significantly more likely to choose the confirmed hypothesis (mean = .59; $SD = .15$; 95% CI = .53-.65) than the (on average equally probable) disconfirmed hypothesis, $\text{Exp}(B) = 1.39$; CI = 1.19-1.64. The proportion of choices favoring the confirmed hypothesis (across all types of

evidence) is displayed in the upper panel of Figure 3. Again, the type of evidence was a significant predictor of participants' choices for the confirmed hypothesis, $F(1, 2,558) = 52.39, p < .001$. Specifically, participants more frequently selected the confirmed hypothesis when the evidence was pattern (mean = .68; $SD = .18$) rather than shape (mean = .49; $SD = .15$). As in Experiment 1, the type of evidence also predicted the RTs, $F(1, 2,558) = 27.57, p < .001$. In particular, RTs were longer when the type of evidence was pattern (mean = 1,389ms; $SD = 448$ ms) rather than shape (mean = 1,195ms; $SD = 354$ ms).

GLMM₄ revealed that, overall, the less probable hypotheses were less likely to be chosen (mean = .39; $SD = .01$; 95% CI = .38-.42) than the more probable ones, $\text{Exp}(B) = .671$; CI = .59-.76. The GLMM₅ analysis showed a significant effect of the agreement between posterior probability and impact on participants' errors, $F(1, 2,046) = 29.04, p < .001$. As in Experiment 1, the number of errors was greater when the more probable hypothesis was disconfirmed (mean = .49; $SD = .18$) than confirmed (mean = .30; $SD = .18$) by the evidence; the proportion of errors in the two conditions for each participant is reported in the lower panel of Figure 3. Again, the agreement between posterior probability and impact did not significantly predict the time required to make a choice, $F(1, 2,046) = .42, p = .519$ (mean = 1,314ms, $SD = 422$ ms and mean = 1,275ms, $SD = 390$ ms, for disagreement and agreement, respectively).

Therefore, the results of Experiment 2 replicated those of Experiment 1 and showed that overall the confirmed hypothesis was selected more often than the (on average equally probable) disconfirmed hypothesis, and that errors increased when there was no agreement between evidential impact and posterior probability. The choice for the confirmed hypothesis, however, was still influenced by the type of evidence: participants more frequently chose the confirmed hypothesis when they were provided with the pattern of the figure and had to predict its shape rather than when it was the other way around. Therefore, yet again, the shape of the figure appeared as the least salient feature, and, exactly as observed in Experiment 1, RTs were

faster when the other feature (in this case pattern) was used as hypothesis rather than evidence. This is compatible with the possibility that evidence about shape was not salient enough to drive a probabilistic updating of the pattern.

Experiment 3

The replacement of color with pattern which was intended to eliminate the asymmetry observed between the two types of evidence employed in Experiment 1 was unsuccessful since pattern also resulted as being more compelling than shape in determining participants' choices. To contrast such asymmetry, in Experiment 3, we employed figures with two different features: line orientation (i.e., horizontal vs. vertical) and line curvature (i.e., wavy vs. straight), which have been claimed to be among the simplest line properties (Treisman & Gormican, 1988) and most relevant visual search attributes (Wolfe & Horowitz, 2017).

Results

As in Experiments 1 and 2, participants were significantly more likely to choose the confirmed hypothesis (mean = .63; $SD = .14$; 95% CI = .58-.68) than the (on average equally probable) disconfirmed hypothesis, $\text{Exp}(B) = 1.75$; CI = 1.39-2.19. The proportion of choices favoring the confirmed hypothesis (across all types of evidence) is displayed in the upper panel of Figure 4. Unlike Experiments 1 and 2, however, the type of evidence was not a significant predictor of participants' choices for the confirmed hypothesis, $F(1, 2,558) = .97$, $p = .326$: participants selected the confirmed hypothesis equally often when the type of evidence was line orientation (mean = .64; $SD = .16$) or line curvature (mean = .61; $SD = .18$). Nevertheless, the type of evidence significantly affected participants' RTs, $F(1, 2,558) = 6.87$, $p = .009$. Specifically, choices took longer when the type of evidence was line curvature (mean = 1,330ms; $SD = 461$ ms) rather than line orientation (mean = 1,245ms; $SD = 453$ ms).

As in previous experiments, GLMM₄ revealed that, overall, the less probable hypotheses were less likely to be chosen (mean = .40; $SD = .01$; 95% CI = .38-.42) than the more

probable ones, $\text{Exp}(B) = .670$; $\text{CI} = .60-.74$. GLMM₅ revealed that the agreement between posterior probability and impact was a significant predictor of participants' errors, $F(1, 2,046) = 39.47$, $p < .001$. Errors were more frequent when the evidence disconfirmed (mean = .53, $SD = .18$) rather than confirmed (mean = .27, $SD = .14$) the more likely hypothesis; the proportion of errors in the two conditions for each participant is reported in the lower panel of Figure 4. In this last experiment, however, the agreement between posterior probability and impact significantly predicted also RTs, $F(1, 2,046) = 6.51$, $p = .011$: participants' choices were faster when posterior probability and impact were in agreement (mean = 1,256ms, $SE = 472$ ms) rather than in disagreement (mean = 1,321ms, $SE = 420$ ms).

To summarize, Experiment 3 replicated one more time the two main effects of interest documented in Experiments 1 and 2: the confirmed hypothesis was selected more often than the (on average equally probable) disconfirmed hypothesis, and errors were significantly more frequent when impact and posterior probability were dissociated. However, in Experiment 3, the type of evidence did not affect participants' choices for the confirmed hypothesis, indicating that the combination of line curvature and orientation was the most balanced pair of features among those employed in the present study (even if RTs were longer for arguments which employed line curvature as evidence). This clean pattern is corroborated further by longer RTs when impact and posterior probability were dissociated, a result that, consistently with the increased number of errors observed in the same condition, suggests a more difficult inference.

Table 4.

Main results of the six GLMM analyses performed for Experiment 1, 2, and 3

Experiment	Analysis	Dependent variable	Predictors	<i>B</i>	SE	<i>t</i>	Sig.	Exp(<i>B</i>)	CI
1	GLMM ₁	Choice ^a	Intercept	.315	.093	3.391	.001	1.370	1.14/1.64
	GLMM ₂	Choice ^a	Intercept	.953	.131	7.270	.000	2.594	2.01/3.35
			Type of evidence ^c	-1.177	.158	-7.467	.000	.308	.27/.42
	GLMM ₃	RT	Intercept	7.54	.049	155.296	.000		7.45/7.64
			Type of evidence ^c	-.097	.023	-4.207	.000		-.14/-.05
	GLMM ₄	Error ^b	Intercept	-.376	.044	-8.577	.000	.686	.63/.75
GLMM ₅	Error ^b	Intercept	-.071	.103	-.687	.492	.931	.76/1.14	
		<i>Pr-Imp</i> agreement ^d	-.662	.148	-4.48	.000	.516	.39/.69	
GLMM ₆	RT	Intercept	7.503	.048	155.803	.000		7.41/7.60	
		<i>Pr-Imp</i> agreement ^d	-.011	.019	-.580	.562		-.05/.03	
2	GLMM ₁	Choice ^a	Intercept	.331	.082	4.040	.000	1.393	1.19/1.64
	GLMM ₂	Choice ^a	Intercept	.714	.100	7.106	.000	2.042	1.68/2.49
			Type of evidence ^c	-.738	.102	-7.238	.000	.478	.39/.58
	GLMM ₃	RT	Intercept	7.06	.057	124.727	.000		6.95/7.17
			Type of evidence ^c	-.147	.028	-5.251	.000		-.20/-.09
	GLMM ₄	Error ^b	Intercept	-.399	.063	-6.283	.000	.671	.59/.76
GLMM ₅	Error ^b	Intercept	-.027	.103	-.268	.789	.973	.80/1.19	
		<i>Pr-Imp</i> agreement ^d	-.797	.148	-5.389	.000	.451	.34/.60	
GLMM ₆	RT	Intercept	6.993	.056	123.678	.000		6.88/7.10	
		<i>Pr-Imp</i> agreement ^d	-.017	.026	-.645	.519		-.07/.034	
3	GLMM ₁	Choice ^a	Intercept	.558	.116	4.786	.000	1.747	1.39/2.19
	GLMM ₂	Choice ^a	Intercept	.653	.143	4.546	.000	1.922	1.45/2.55
			Type of evidence ^c	-.157	.159	-.983	.326	.855	.63/1.17
	GLMM ₃	RT	Intercept	6.945	.048	143.733	.000		6.85/7.04
			Type of evidence ^c	.080	.031	2.621	.009		.02/.14
	GLMM ₄	Error ^b	Intercept	-.401	.054	-7.456	.000	.670	.60/.74
GLMM ₅	Error ^b	Intercept	.138	.131	1.056	.291	1.148	.89/1.48	
		<i>Pr-Imp</i> agreement ^d	-1.183	.188	-6.282	.000	.306	.21/.44	
GLMM ₆	RT	Intercept	7.014	.047	147.887	.000		8.92/7.11	
		<i>Pr-Imp</i> agreement ^d	-.056	.022	-2.552	.011		-.10/-.01	

^a Choice is codified as 1 = choice for the confirmed alternative and 0 = choice for the disconfirmed alternative^b Error is codified as 0 = choice for the more probable alternative; 1 = choice for the less probable alternative^c Type of evidence is codified as 1 = pattern and 2 = shape^d *Pr-Imp* agreement is codified as 0 = disagreement between posterior probability and impact and 1 = agreement between posterior probability and impact

Discussion

Inductive inferences are being drawn continuously. Whether voluntary or automatic, they all contribute to an active understanding of the world and determine how efficiently we adapt to it. Improving knowledge of the extent to which such inferences are accurate and, possibly, identifying their biasing factors is important not only for theoretical reasons but also because it may allow us to prevent the most common mistakes from taking place. The three experiments reported in this study provided consistent data (both across individuals and across stimuli) indicating that probabilistic judgments concerning simple visual stimuli are systematically distorted by impact relations between evidence and hypothesis. Indeed, all things being equal, participants tended to overestimate [underestimate] the probability of confirmed [disconfirmed] hypotheses and made more errors in their probability judgments when the hypotheses were disconfirmed by the evidence (i.e., when posteriors and impacts were dissociated).¹

The results of this study complemented and extended those obtained in previous works which employed arguments whose evaluation involved reference to a large body of beliefs and assumptions (such as, for example, the common knowledge about gender differences in interests or hobbies). They supported the centrality of the detection of impact relationships for inductive reasoning and reinforced the conjecture that, although impact is formally defined as a function of probability values, psychologically it constitutes a primitive kind of judgment that people implicitly make when reasoning under uncertainty, even in the case of purely probabilistic tasks (for a similar conclusion in the context of the conjunction fallacy, see Crupi et al.,

¹ Note that, as pointed out by an anonymous reviewer, the results of this study are compatible with the occurrence of a base-rate fallacy. Unlike classic base-rate fallacy scenarios, participants were not explicitly provided with numerical information regarding base rates nor regarding the statistical association between evidence and hypotheses. However, when posteriors and impacts were dissociated, incorrect responses mathematically implied that base rates were not appropriately considered. Such a connection between the effect of evidential impact on posterior probability judgments and the base-rate fallacy is not surprising since, as mentioned in the Introduction, the most convincing explanations of the latter (as well as of other probabilistic reasoning errors) make use of concepts such as representativeness or relevance that can be (and have been) formalized in terms of a prevalence of impact over probability judgments.

2007, and Tentori et al., 2013). Moreover, for the very first time, our results showed that the effect of impact on probability judgments extends to perceptual materials and, in particular, to arguments that concern arbitrary artificial associations between evidence and hypothesis. Such a finding reveals that the same inferential processes that bias probabilistic reasoning, language and associative learning may also affect more basic cognitive functions, as perception. Furthermore, it casts serious doubts on the possibility that the preeminence of impact over probability judgments has an intrinsic epistemological value. Indeed, while the direction of impact relationships between real-world evidence and hypotheses typically conveys important information, for example, with regard to the existence and nature of causal links, in case of simple perceptual arguments like those used in these experiments, impact relations express a purely structural property of the stimuli, which is completely uninformative for the probability task considered.

The results of this study might also contribute to the growing literature aimed at understanding how the perceptual system encodes information of visual stimuli, since they specify a new and well-determined way in which the internal representation of the association between basic features and consequent inferences might be sub-optimal. Indeed, we can be quite confident that our participants' probability errors did not stem from the difficulty of interpreting the sensory data, because the perceptual features employed in the stimuli were simple and distinguishable, and, especially for Experiments 2 and 3, time was more than enough to fully explore the set of figures. It is also unlikely that errors originated from an ambiguity of priors, since, in such an artificial context, these were completely independent of any pre-existing background knowledge and corresponded exactly to the (available) base-rates of the two alternative hypotheses. Our results showed that the observed probability errors were not random either but depended on one specific aspect of the relationship between hypotheses and evidence: the direction of impact. To our knowledge, such a source of bias has not yet been considered within

the literature on perceptual updating and could offer a new perspective on Bayesian inference in this field.

Finally, the asymmetry in the effect of impact found in Experiments 1 and 2 seems to indicate that, all probabilities being equal, the inductive inferences between visual features of different salience differed depending on whether the most salient feature was used as evidence or hypothesis. More specifically, our results suggest that participants used information on the most salient feature to update their predictions about the less salient feature, but not vice-versa. When the evidence concerned the less salient feature, in fact, participants' judgments seemed driven by the priors of the values of the most salient feature, without these being updated on information regarding the less salient feature. If confirmed in further studies, such a pattern would indicate that inferences involving visual features with varying levels of salience depend mainly on the information conveyed by the most salient feature (for a similar conclusion in the domain of the detection of regularities from multiple feature dimensions, see Yu, Luo, Osherson & Zhao, 2019).

The very limited range of impact (around $\pm .4$ or $\pm .3$) and posterior probability (.4-.5, .45-.55) values in our stimuli, on one hand, allowed us to fully balance these two variables while keeping the overall number of figures rather low, but on the other, did not permit us to insert these variables as quantitative predictors into our analyses. Future work might then consider a refinement of our experimental design by quantifying more precisely the distortion in percepts that can be ascribed to various degrees of impact relations across various levels of posteriors.

Another investigation that, in our opinion, is worth pursuing in future research concerns why people's probability judgments in various reasoning, language and perception tasks are systematically biased by impact relationships, and more in general why there is a prominence of evidential over posterior probability assessment. The answer to these questions constitutes

an intriguing puzzle from an evolutionary point of view since, as specified in the Introduction section, probabilistic reasoning is acknowledged as crucial for many important activities, such as decision making between risky prospects. According to what has been conjectured by Tenenori, Chater, and Crupi (2016), people might be more sensitive to impact relationships than they are to posterior probabilities because, by tracking direct or indirect causal dependencies between evidence and hypothesis, the former achieves a higher degree of stability than the latter in response to background knowledge or changing environments. Such a proposal is inspired by the assumption that confirmation is a necessary (albeit not sufficient) condition for probabilistic causation and considers the ability to detect impact relationships as fundamental for building reliable and stable representations of the world. The replication of the effect of impact on probability judgments concerning artificial perceptual materials documented in this study offers a different, although potentially complementary, perspective. The notion of Bayesian confirmation has a striking formal correspondence with that of Bayesian surprise, which has been defined exactly in terms of some divergence between posterior and prior beliefs (for two proposals see, for example, Itti & Baldi, 2009, and Reisenzein, Horstmann, & Schützwohl, 2019). Since surprise is acknowledged as one of the main attractors of human attention and a driver for learning, memory, and creativity (Ranganath & Rainer, 2003, Macedo & Cardoso, 2019), one might speculate that prioritizing evidential assessment is well-suited to detecting novelty in incoming data and that, in the long run, the epistemic advantage provided by this surprise-search behavior well compensates for suboptimal accuracy of probability estimates. Hopefully, new experimental paradigms and simulations aimed at estimating more directly the costs and benefits of evidence versus probability evaluation across various cognitive domains will allow us a more in-depth exploration of these, and maybe other, suggestions.

Compliance with Ethical Standards

Funding: The work was partially funded by the European Research Council (ERC Consolidator Grant 617629, awarded to Giorgio Coricelli).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethical approval: The study was approved by the Human Research Ethics committee of the University of Trento.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Data Deposition Information: Experimental stimuli and data sets have been deposited in the Open Science Framework, and are publicly available at doi: 10.17605/OSF.IO/QNEMY (<https://osf.io/qnemy/>)

References

- Adamo, M., Wozny, S., Pratt, J., & Ferber, S. (2010). Parallel, independent attentional control settings for colors and shapes. *Attention, Perception, & Psychophysics*, *72*, 1730-1735. <https://doi.org/10.3758/APP.72.7.1730>
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*, 211–233. [https://doi.org/10.1016/0001-6918\(80\)90046-3](https://doi.org/10.1016/0001-6918(80)90046-3)
- Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology*, *65*, 1119-1131. <http://dx.doi.org/10.1037/0022-3514.65.6.1119>
- Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*, 1-20. <http://dx.doi.org/10.1037/rev0000047>
- Breitmeyer, B. G., Ogmen, H., & Chen, J. (2004). Unconscious priming by color and form: Different processes and levels. *Consciousness and Cognition: An International Journal*, *13*, 138-157. <http://dx.doi.org/10.1016/j.concog.2003.07.004>
- Bullinaria, J.A., & Levy, J.P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*, 510-526. <https://doi.org/10.3758/BF03193020>
- Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Cognitive Science*, *1*, 811-823. <https://doi.org/10.1002/wcs.79>
- Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, *74*, 229–252. <https://doi.org/10.1086/520779>
- Crupi, V., Tentori, K., & Lombardi, L. (2009). Pseudodiagnosticity revisited. *Psychological Review*, *116*, 971-985. <http://dx.doi.org/10.1037/a0017050>
- Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.),

Formal Representation of Human Judgment (pp. 17–52). New York: Wiley.

- Fitelson, B. (1999). The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66, S362–S378. <https://doi.org/10.1086/392738>
- Good, I.J. (1984). The best explicatum for weight of evidence. *Journal of Statistical Computation and Simulation*, 19, 294–9. <https://doi.org/10.1080/00949658408810739>
- Griffiths, T.L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are: Comment on Bowers and Davis (2012)). *Psychological Bulletin*, 138, 415-422. <https://psycnet.apa.org/doi/10.1037/a0026884>
- Heit, E., & Hahn, U. (2001). Diversity-based reasoning in children. *Cognitive Psychology*, 43, 243-273. <https://doi.org/10.1006/cogp.2001.0757>
- Hogarth, R.M., & Einhorn, H.J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24, 1-55. [https://doi.org/10.1016/0010-0285\(92\)90002-J](https://doi.org/10.1016/0010-0285(92)90002-J)
- Itti, L. & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*. 49, 1295-1306. <https://doi.org/10.1016/j.visres.2008.09.007>
- Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251. <https://doi.org/10.1037/h0034747>
- Kemeny, J. & Oppenheim, P. (1952). Degrees of factual support. *Philosophy of Science*, 19, 307–24.
- Kemp, C., & Tenenbaum, J.B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, 116, 20-58. <http://dx.doi.org/10.1037/a0014282>
- Körding, K.P., & Wolpert, D.M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, 10, 319-326. <https://doi.org/10.1016/j.tics.2006.05.003>
- Lo, Y., Sides, A., Rozelle, J., & Osherson, D. (2002). Evidential diversity and premise probability in young children's inductive judgment. *Cognitive Science*, 26, 181-206.

https://doi.org/10.1207/s15516709cog2602_2

López, A., Gelman, S.A., Gutheil, G., & Smith, E.E. (1992). The development of category-based induction. *Child Development*, 63, 1070-1090.

<http://dx.doi.org/10.2307/1131519>

Macedo, L. & Cardoso, A. (2019). A contrast-based computational model of surprise and its applications. *Topics in Cognitive Science*, 11, 88-102.

<https://doi.org/10.1111/tops.12310>

Marcus, G.F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, 24, 2351-2360.

<https://doi.org/10.1177/0956797613495418>

Mastropasqua, T., Crupi, V., & Tentori, K. (2010). Broadening the study of inductive reasoning: Confirmation judgments with uncertain evidence. *Memory & Cognition*, 38, 941-

950. <https://doi.org/10.3758/MC.38.7.941>

Nadalini, A., Marelli, M., Bottini, R., and Crepaldi, D. (2018). Local associations and semantic ties in overt and masked semantic priming. *Proceedings of the Fifth Italian Conference on Computational Linguistics*. Torino.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive skills and their acquisition*, 1(1981), 1-55.

Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97, 185-200. [http://dx.doi.org/10.1037/0033-](http://dx.doi.org/10.1037/0033-295X.97.2.185)

[295X.97.2.185](http://dx.doi.org/10.1037/0033-295X.97.2.185)

Paperno, D., Marelli, M., Tentori, K., & Baroni, M. (2014). Corpus-based estimates of word association predict biases in judgment of word co-occurrence likelihood. *Cognitive Psychology*, 74, 66-83. <https://doi.org/10.1016/j.cogpsych.2014.07.001>

Ranganath C. & Rainer G. (2003). Neural mechanisms for detecting and remembering novel

- events. *Nature Reviews in Neuroscience*, 4, 193–202. <https://doi.org/10.1038/nrn1052>
- Reisenzein, R., Horstmann, G., & Schützwohl, A. (2019). The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science*, 11, 50–74. <https://doi.org/10.1111/tops.12292>
- Rescorla, M. (2015). Bayesian perceptual psychology. In M. Matthen (Ed.), *The Oxford Handbook of the Philosophy of Perception* (pp. 694-716). New York: Oxford University Press. <https://dx.doi.org/10.1093/oxfordhb/9780199600472.013.010>
- Rusconi, P., Marelli, M., D'Addario, M., Russo, S., & Cherubini, P. (2014). Evidence evaluation: Measure Z corresponds to human utility judgments better than measure L and optimal-experimental-design models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40, 703-723. <https://doi.org/10.1037/a0035549>
- Shen, S., & Ma, W. J. (2016). A detailed comparison of optimality and simplicity in perceptual decision making. *Psychological Review*, 123, 452-480. <http://dx.doi.org/10.1037/rev0000028>
- Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, 124, 410-441. <https://psycnet.apa.org/doi/10.1037/rev0000052>
- Tenenbaum, J.B., Griffiths, T.L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, 10, 309-318. <https://doi.org/10.1016/j.tics.2006.05.009>
- Tentori, K. & Crupi, V. (2012). How the conjunction fallacy is tied to probabilistic confirmation: Some remarks on Schupbach (2009). *Synthese*, 184, 3-12. <http://dx.doi.org/10.1007/s11229-009-9701-y>
- Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007). Comparison of confirmation measures. *Cognition*, 103, 107-119. <http://dx.doi.org/10.1016/j.cognition.2005.09.006>

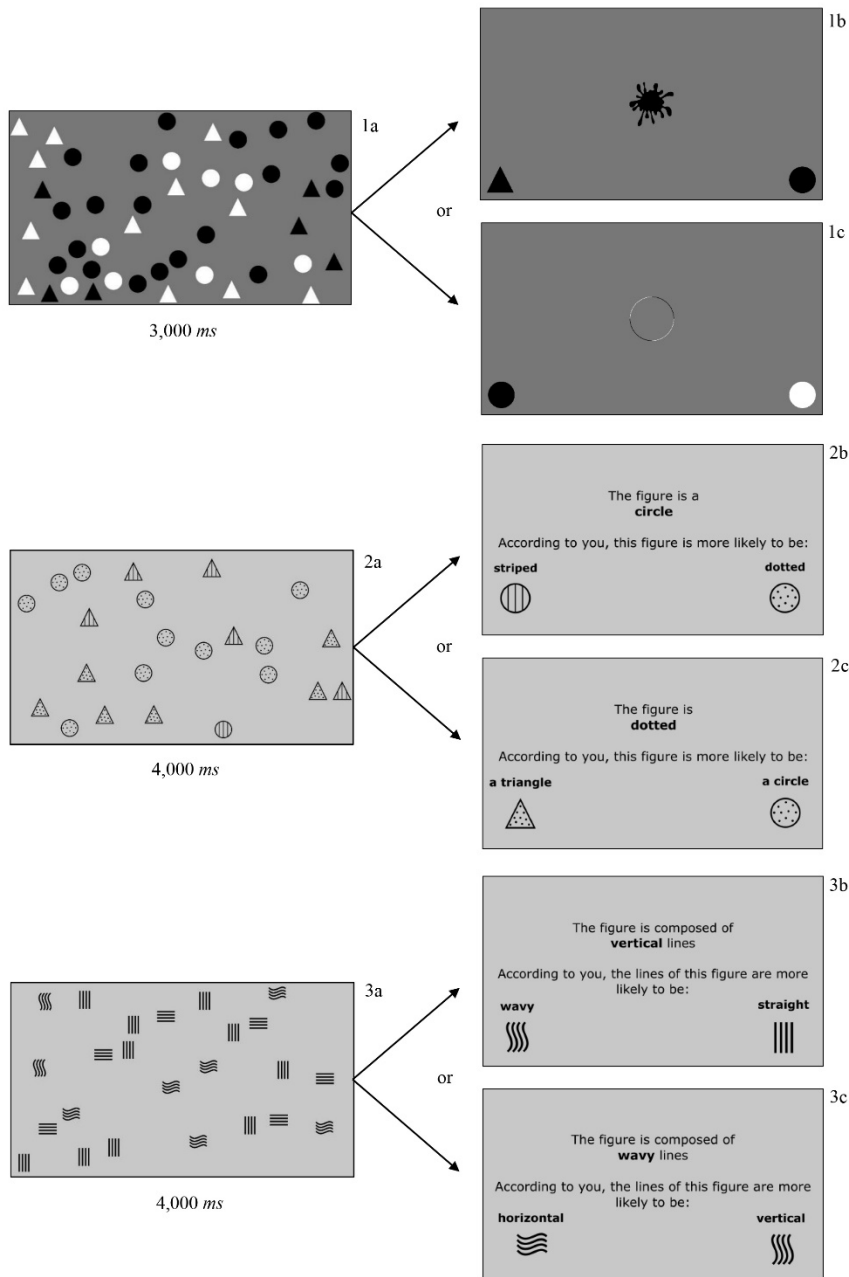
- Tentori, K., Chater, N., & Crupi, V. (2016). Judging the probability of hypotheses versus the impact of evidence: Which form of inductive inference is more accurate and time-consistent? *Cognitive Science*, *40*, 758-778. <https://doi.org/10.1111/cogs.12259>
- Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*, 235-255. <http://dx.doi.org/10.1037/a0028770>
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological Review*, *95*, 15-48. <http://dx.doi.org/10.1037/0033-295X.95.1.15>
- Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 153–160). New York: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90*, 293-315. <http://dx.doi.org/10.1037/0033-295X.90.4.293>
- Turpin, M.H., Meyers, E.A., Walker, A.C., Białek, M., Stolz, J.A., & Fugelsang, J.A. (2020). The environmental malleability of base-rate neglect. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-020-01710-1>
- Wolfe, J.M., & Horowitz, T.S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 495-501. <https://doi.org/10.1038/nrn1411>
- Wolfe, J.M., & Horowitz, T.S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*, 0058. <https://doi.org/10.1038/s41562-017-0058>
- Yu, R.Q., Luo, Y., Osherson, D., & Zhao, J. (2019). Perception of multi-dimensional regularities is driven by salience. *Attention, Perception, & Psychophysics*, *81*, 1564–1578.

<https://doi.org/10.3758/s13414-019-01667-x>

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*, 301-308. <https://doi.org/10.1016/j.tics.2006.05.002>

Zhong, L., Lee, M., Huang, Y., & Mo, L. (2014). Diversity effect in category-based inductive reasoning of young children: Evidence from two methods. *Psychological Reports*, *114*, 198–215. <http://dx.doi.org/10.2466/10.11.pr0.114k17w3>

Figure 1. Examples of sets of figures used in the three experiments, with corresponding questions and answers.



In each trial, participants were first presented with a set of figures on uniform grey background. Figures were made up of two features with two possible values each (i.e., black or white color and circular or triangular shape in Exp. 1, striped or dotted pattern and circular or triangular shape in Exp. 2, horizontal or vertical line orientation and wavy or straight curvature in Exp. 3, for examples, see 1a, 2a, and 3a respectively). At the end of the presentation time, a figure was drawn from the set and the value of one of its features was revealed (i.e., “black” in 1b). In light of this evidence, participants had to report their expectation about the value of the other feature of the figure by selecting one of the two icons at the bottom of the screen (which counted as alternative hypotheses, i.e., “circle” vs. “triangle” in 1b). In Exp 2 and 3, we communicated the evidence verbally to prevent the graphical display of the figure from potentially cueing the levels of the other feature. Once the response was provided, the selected hypothesis appeared in the center of the screen for 1.5s, and participants were prompted to press the spacebar to proceed with the next trial.

Figure 2. For each participant of Experiment 1, the upper panel displays the proportion of choices favoring the confirmed hypothesis (across all types of evidence), while the lower panel displays the proportion of errors when posterior probability and impact were in agreement versus disagreement (and related group means).

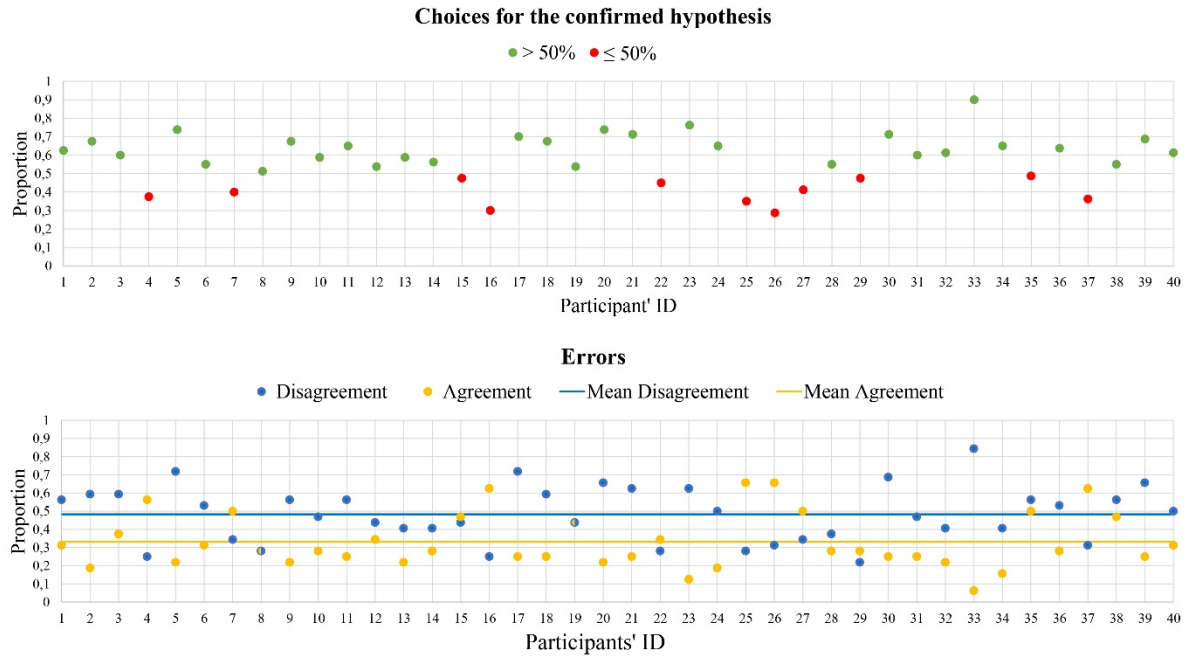


Figure 3. For each participant of Experiment 2, the upper panel displays the proportion of choices favoring the confirmed hypothesis (across all types of evidence), while the lower panels displays the proportion of errors when posterior probability and impact were in agreement versus disagreement (and related group means).

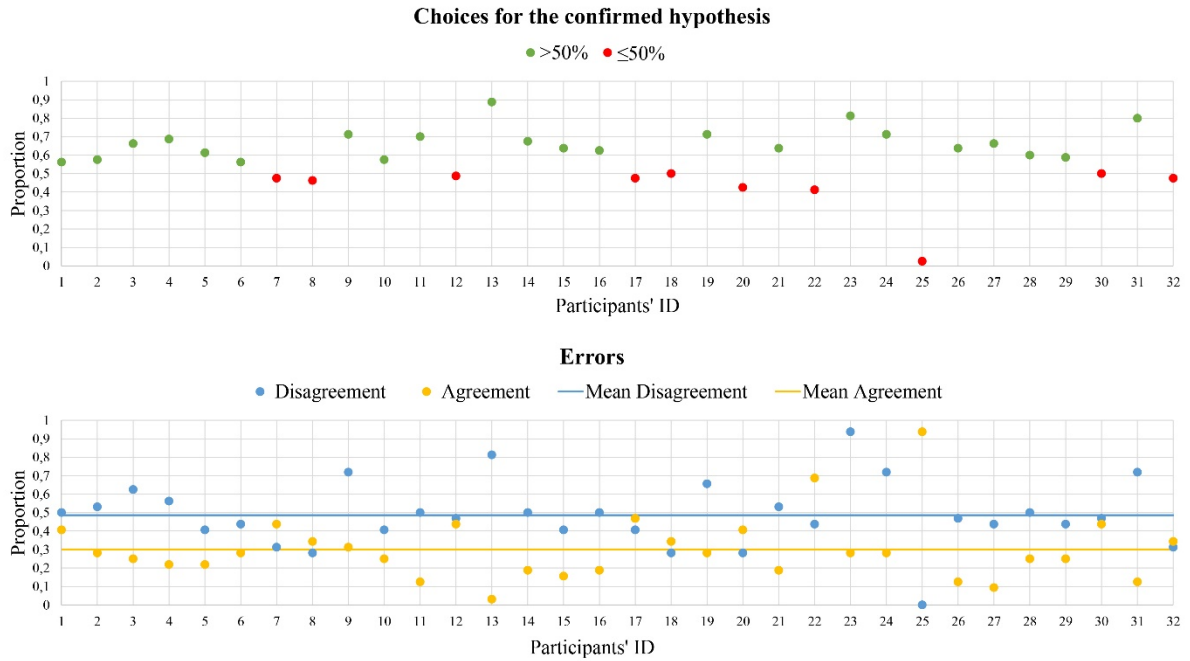


Figure 4. For each participant of Experiment 3, the upper panel displays the proportion of choices favoring the confirmed hypothesis (across all types of evidence), while the lower panels displays the proportion of errors when posterior probability and impact were in agreement versus disagreement (and related group means).

