UNIVERSITY OF TRENTO - Italy

CIMeC – Centre for Mind/Brain Sciences

Doctoral School in Cognitive and Brain Sciences (XXXI Cycle)

Track Cognitive Neuroscience

THE EFFECT OF EVIDENTIAL IMPACT ON PERCEPTUAL PROBABILISTIC REASONING

Doctoral candidate: Marta Mangiarulo

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive and Brain Sciences

April 2019

Table of Contents

List of tables

List of figures

Abstract

For decades, works in psychology of thinking and decision making have been reporting suboptimal performance and systematic departures from the axioms of probability theory in people's probability judgments. In these first works, poor performance was often attributed to people making normatively wrong intuitions because of their limited cognitive resources and lack of statistical skills. Over the last years, studies that considered various Bayesian models of inductive reasoning but also other high and lower-level cognitive processes provided a more optimistic picture by showing that, despite departing from the normative benchmark, people's reasoning skills lead to adaptive and sound performance in everyday life. Different explanatory accounts for this suboptimal but sound reasoning have been proposed, some being more compelling than others. The present thesis is aimed at exploring one of these accounts that is based on confirmation relations and suggests that human inductive ability might rely more on estimating evidential impact than posterior probability. So far, this account has been applied to classical probabilistic reasoning errors, linguistic and psycholinguistic phenomena and probabilistic inferences with verbal stimuli. In this study, we tried to see whether the implicit estimation of confirmation relations can affect probability judgments also when the link between evidence and hypotheses is operationalized as the arbitrary association between visual features in briefly presented figures. First, we expected participants to consider confirmed hypotheses more probable than corresponding (in terms of posterior probability) disconfirmed ones; second, we expected them to choose the more likely option (i.e. the normatively correct one) more often when it was confirmed by the evidence provided than when it was disconfirmed. Four computer-based experiments were conducted using the same methodology. Experimental stimuli consisted of inductive arguments concerning 40 sets of figures composed of two features with two possible values each. By varying the probabilistic association between the two values of the features, sets were generated to have, for each possible combination of the two features, two arguments with the same posteriors and opposite impacts. In each trial, participants first looked at a set of figures. One of these figures was then randomly drawn. Participants were informed about the value of one feature of the drawn figure (e.g., that it was a "circle") and had to guess the value of the other feature ("white" vs. "black").

Throughout the four experiments, we used three different combinations of features: color and shape (exp.1: black/white; exp 2: light/dark grey), pattern and shape (exp 3) and type and orientation of line (exp 4).

In all four experiments, participants systematically chose the confirmed alternative over the equally probable, but disconfirmed one, and chose the normatively incorrect (i.e. less likely) alternative more often when it was confirmed (vs. disconfirmed) by the evidence provided. These results provided a first empirical evidence of the effect of confirmation relations on probability judgment with perceptual stimuli, but also highlighted a significant influence of the experimental material itself on choice patterns. In fact, in experiments 1 to 3 the obtained results showed that color (or pattern) was a more compelling evidence than shape in determining participants' choices. The combination of line curvature and orientation used in experiment 4 proved to be the more balanced among those employed in the present research. Only in this last experiment, indeed, the type of evidence did not affect the choice for the confirmed alternative, nor the amount of errors. The results we found supported our experimental claims showing that confirmation relations can affect probability judgments even in absence of any semantic element, but also suggested the existence of a mutual influence between perceptual features and probability judgments. Our experimental results have theoretical as well as applied implications. On a theoretical level, they extend the results coming from works involving verbal and linguistic material to perceptual stimuli with no semantic background. Additionally, they show that high-level relations, which are completely unknown to the subject, affect the way people perceive relations within a visual set of perceptual items. This might have interesting and noteworthy implications for studies on visual cognition, and, on a broader level, contingency learning and stereotypical judgments.

Chapter 1: Introduction

This work is aimed at investigating whether and how confirmation relations affect probabilistic inferences with perceptual material. Evidence of the effect of confirmation relations on probability judgments has already been found in the literature (Paperno et al., 2014; Tentori, Chater and Crupi, 2016) in works involving verbal stimuli and linguistic corpora. However, no evidence of this effect on probabilistic reasoning with perceptual material has been found so far. In order to tackle this issue, the present work involves concepts and experimental methodologies coming from epistemology, psychology and psychophysics. In Chapter 2, I will introduce and compare contributions from epistemology and psychology to the understanding of probabilistic reasoning; the first part will discuss psychological approaches to probability judgment and the second one will present confirmation measures and their role in Bayesian epistemology. In the last section of this chapter, I will present works aimed at comparing directly these two concepts. In Chapter 3, I will discuss an issue that I believe would benefit from a multidisciplinary investigation involving philosophy, psychology and psychophysics: the Bayesian brain hypothesis. In Chapter 4, I will present four behavioral experiments aimed at testing the hypotheses we derived from the literature discussed, and finally, in Chapter 5, I will consider our experimental results in light of the existing literature and discuss the strengths and limitations of the present study, as well as further developments and theoretical or practical implications for other research domains.

1.1 Theoretical background

When exploring the surrounding environment, people generate hypotheses and collect information to test them. This process involves deductive as well as inductive reasoning: people extract hypotheses and rules through induction, then they combine inductive inferences and deduction to create anticipations on the environment based on these rules (see Cherubini, 2005). Deductive reasoning can be defined as that type of reasoning which, given true premises, yields conclusions that are necessarily true. Both inductive and deductive reasoning increase knowledge and provide new information, but new empirical knowledge is only acquired by means of inductive reasoning, which goes beyond the information given to draw novel conclusions. Such conclusions are probable but not logically implied in the already existing evidence: this new knowledge, then, is uncertain. In light of this, the main goal of a theory of inductive reasoning would be to define when some inference is strong or not, for example assigning a certain value to it.

Any inductive inference concerns the relation between two conceptually related but dissociable elements: the hypothesis of interest (h) and the available evidence (e) (Tentori, Chater, & Crupi, 2016). When evaluating an inductive argument, one can focus on the **hypothesis** and ask how probable it is in light of the evidence; this is the posterior probability of h given e and Bayes rule represents its normative benchmark. One can also choose to focus on the **evidence** and ask how much the evidence e increases/decreases the belief in h; this is the impact, or degree of confirmation, of e over h and, in the Bayesian framework, it is quantified by confirmation measures. In the epistemological literature, confirmation measures are also referred to as measures of evidential support and in this work I will use the two terms interchangeably. Evidential support is a relative notion measuring the change in someone's belief; posterior probability, on the other hand, represents its absolute counterpart. These two elements are usually associated in everyday life: if the probability of a certain hypothesis *h* is high given a piece of evidence *e*, conversely it is also likely that the evidential impact of *e* on the probability of *h* will be high. However, they can be disentangled: in experimental settings, it is possible to construct inferential arguments composed of a hypothesis *h* and an evidence *e* where the probability of the hypothesis is high in light of the evidence, but the latter disconfirms the former, or vice versa. According to some recent explanatory accounts of reasoning biases (e.g. Crupi, Fitelson, & Tentori, 2008), this was the case in Kahneman and Tversky's first investigations of probabilistic reasoning (Kahneman & Tversky, 1982, Tversky & Kahneman, 1974). These first experiments involved probabilistic scenarios in which the probability-driven hypothesis, considered the normatively correct one, and the confirmation-driven one would always diverge; as just argued, this is unlikely in most everyday situations. Tversky & Kahneman's (1974) explanation for their findings was that people are poor statistical reasoners and that their intuitions are normatively wrong. More precisely, they claimed that when making judgments under uncertainty, people employ heuristics principles, "which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (Tversky & Kahneman, 1974). However, these experimental findings are counterbalanced by evidence of sound probabilistic reasoning in more naturalistic settings suggesting that maybe there is some element in the experimental task affecting reasoning performances. Tentori, Chater and Crupi (2016)'s confirmation-based account suggested that participants' performance in inferential tasks is apparently suboptimal because it is driven by confirmation relations that make it depart from the normative benchmark. This crucial distinction between probability and impact judgment has been discussed only recently; in the past, in psychology and cognitive science, literature has mainly focused on the assessment of posterior probability, giving much less attention to the estimation of evidential impact except for some recent experimental works (Crupi et al., 2008; Tentori et al., 2016; Tentori, Crupi, & Russo, 2013).

Chapter 2: Psychological and philosophical approaches to probabilistic reasoning

2.1. Probabilistic reasoning: psychological approach

In psychology, experimental investigation of inductive reasoning mainly focused on **posterior probability.** Posterior probability measures the probability of a given hypothesis in light of some information and it can be estimated by means of Bayes rule, which is a "formula that allows us to compute change in belief in a set of hypotheses, H, as a result of encountering some information, D (data)" (Manktelow, 2012). One possible formulation of the rule in formal terms is the following (ivi):

$$p(H|D) = \frac{p(D|H)x\, p(H)}{p(D)}.$$

The *prior knowledge* that hypothesis *H* is true is combined with the *likelihood* of the evidence to derive an estimate of the *posterior probability* that hypothesis *H* is true given the evidence *D*; to perform normatively correct judgments, people must know all these values and be able to combine them in the right way. For years, literature on probabilistic reasoning has shown that people are not capable of optimal probabilistic reasoning in experimental setting (Kahneman & Tversky, 1973, 1982, Tversky & Kahneman, 1974) and that they are subjected to so-called reasoning biases. However, there is a discrepancy between these classical studies and more recent research reporting sound probabilistic performance with naturalistic material (see Domurat et al., 2015; Griffiths, Kemp, & Tenenbaum, 2008; Griffiths & Tenenbaum, 2006; Kemp & Tenenbaum, 2009; Zhao, Shah, & Osherson, 2009).

2.1.1 Recent accounts of probabilistic reasoning

In standard treatments of probability, the conditional probability of a hypothesis given an evidence $Pr(H|D)$ is defined as the ratio of $Pr(A \cap B)$ to $Pr(B)$ (Zhao et al., 2009). However, in everyday life, some of the estimates that Bayes theorem requires might be unavailable, or the computations it entails could be too complicated to perform, leading to apparently suboptimal performance. Despite this, people's probabilistic reasoning usually leads to optimal and sound behavior. Literature on probabilistic reasoning described and discussed several models of apparent suboptimality which still lead to sound inference, each entailing different computations and notions. To understand whether the mind implicitly computes conditional probability by means of the normatively correct equation, Zhao, Shah, & Osherson (2009) investigated the provenance of judgments of conditional probability and provided evidence for non-normative reasoning. Their work involved three behavioral experiments aimed at understanding whether estimates of $Pr(A|B)$ (conditional probability) come from an implicit calculation of $Pr(A \cap B)/Pr(B)$, which would be the normative benchmark, or they derive from other, simpler computations. Experiments 1 and 2 employed the same stimuli: in these experiments, participants were shown a number of sets of geometric shapes on a computer screen; each set was composed of a certain number of triangles, circles and squares in blue, red and green. In Experiment 1, after seeing each set of stimuli, participants had to answer questions about prior and conditional probabilities in the set, both presented as probability questions and as frequency questions. These questions probed Pr(B), $Pr(A \cap B)$ and $Pr(A|B)$. Experiment 2 involved the same experimental sets and questions, as well as an additional question that queried $Pr(\neg A \cap B)$. Experiment 3 was implemented on Amazon Mechanical Turk, involved scenarios describing future events and the four queries introduced in Experiments 1 and 2 ($Pr(B)$, $Pr(A \cap B)$, $Pr(A|B)$, and $Pr(\neg A \cap B)$. Each participant was invited to supply his personal probability for each of the statements they were shown. The first experiment showed that conditional probability does not come from the normatively correct rule, while Experiment 2 suggested an equation that more accurately described it. Finally, Experiment 3 evaluated the same hypotheses tested in Experiments 1 and 2 in the context of subjective estimates of future events. Overall, then, the three studies showed that probability judgments do not originate from mental computations entailed by the standard definition of posterior probability involving Bayes rule and provided an alternative account of the computations people perform when asked to estimate conditional probability. The relevance of this study for our goals is twofold. First, it suggests that standard computations of probability do not account for actual judgments of conditional probability; second, the experimental task involves perceptual material (i.e. bidimensional grids of figures defined by color and shape), which deviates from usual stimuli involved in thinking and reasoning experiments (inductive arguments in verbal form) and does not provide any kind of semantic information.

An alternative explanation for suboptimal reasoning entailed linear additive instead of multiplicative integration of priors and evidence. Juslin, Nilsson, & Winman (2009) suggested that 'normative' probability judgments are not a necessary requisite for rational decisions and that violations of probability theory do not imply suboptimal decision making (i.e. cognitive biases like conjunction error or base rate neglect). They carried out 5 simulations focused, respectively, on the extent to which estimates generated by a model implementing weighted linear additive integration and a model combining noisy probabilities by applying the conjunction rule of probability theory correspond with true conjunctive probabilities (Study 1), on exploring if the results from Simulation Study 1 held also with a more general conceptualization of the random error in the judgment process (Study 2), on understanding whether good prediction of the criterion with suboptimal weights is also possible with linear additive weighting of probabilities (Study 3), on testing how well four models describe decisions based on conjunctive probabilities (Study 4) and on whether Bayes' theorem is a better tool for using noisy probabilities to estimate a true posterior probability than linear additive integration (Study 5). Taken together, the studies showed that reasoning based on approximate rather than exact knowledge of probabilities and linear additive integration rather than multiplicative represent satisfying solutions in everyday life and they can be as accurate as estimates based on probability theory. The authors discussed the reported results suggesting that people could be more prone to using weighted additive rather than multiplicative integration of information, which can lead to reasoning biases. This finding shows that despite the lack of possibility or incentive to align to probability rules in everyday life, people's reasoning might be still efficient thanks to linear additive integration. However, this strategy deviates from the normative benchmark, i.e., probability theory, therefore is normatively suboptimal.

In a more recent work involving a novel estimation task, Acerbi, Vijayakumar, & Wolpert (2014) investigated how people combine sensory information with statistics collected from prior experience to yield more or less optimal behavior and, more precisely, how priors affect Bayesian computations with explicitly provided probabilistic information. To do so, a behavioral estimation task was used to probe the sources of suboptimality in probabilistic inference; given a probabilistic cue about its location, participants had to locate a specific, unknown target from a set of potential ones. Information consisted of a visual representation of the a priori probability distribution of targets for that trial and a noisy cue about the actual target position. A model-free analysis was performed on participants' performance and showed that it was suboptimal; next, a model with a factorial approach was applied to the data. This model consisted of the combination of four basic factors: decision making, cue-estimation sensory noise, noisy estimation of the prior, and lapse. The results showed that participants' performance was qualitatively in line with Bayesian rules, although suboptimal; this suboptimality was relatively independent from the priors and level of noise in the cue but strongly affected by the class of distribution (i.e. the specific shape of the prior). Subjects' performance, then, was driven by a combination of (noisy) estimation of the parameters of the priors and noisy posteriors and would not align with models of stochastic behavior like probability matching or sample-averaging.

By means of two behavioral experiments and an analytical study, Domurat et al. (2015) provided further evidence for suboptimal but satisfactory probabilistic reasoning in everyday life. They investigated how often people's probability estimates conform to Bayes rule when natural sampling is involved and showed that using Bayes' rule is not necessary to make choices that satisfy it. Experiments 1 and 2 involved 16 computer tasks, composed of a learning stage and a choice stage. During the learning stage, participants could gather information about the environment and develop hypotheses about probabilistic relationship. Three kinds of information could be learned through natural sampling: prior occurrences, likelihood ratios and Bayesian estimates (conditional probabilities). Then, participants would have to use this information in the following phase to carry out choices conforming to Bayes rule. While experiment 1 relied on computer-based tasks, experiment 2 also collected participants' verbal protocols. Finally, study 3 exclusively relied on computer simulations. The goal of this final experiment was to investigate how often strategies causing fallacious responses (*representativeness*, *evidence-only*, *pre-Bayesian*) lead to the same response as Bayes rule. This experiment involved computer simulations of the probabilistic scenarios from studies 1 and 2; all combinations of the information that could be gathered (priors, likelihoods) were generated, in order to explore whether no Bayesian-strategies could provide results as good as the normative ones. The authors analyzed the simulated strategies with regard to (1) different frequencies expressing decision-makers' natural sampling experiences and (2) different base rates, arbitrarily defined as rare [$P(H) \leq 0.25$], frequent [$P(H) \geq 0.75$], and medium [$0.25 < P(H) < 0.75$]. The analysis of these simulated scenarios showed that representativeness and evidence-only strategies lead to choices in line with Bayes rule if base rates are high and the natural sampling size is low. This means that, under specific circumstances and elementary situations, even heuristic strategies can handle probabilistic tasks effectively (as already pointed out by Tversky & Kahneman, 1974). Overall, this paper showed that Bayesian inference could be unnecessary in making correct choices in elementary situations through natural sampling.

Evidence supporting the efficacy of an overall 'suboptimal' strategy also came from Laquitaine & Gardner (2018), who tested whether Bayesian processes can explain people's behavior in a motion direction estimation and a spatial orientation tasks. Several different models were proposed to describe human behavior and compared to the Basic Bayesian one, involving optimal integration between priors and sensory likelihoods. The results showed that one of the alternative models, the Switching observer, described subjects' estimates better than the Basic Bayesian observer, indeed providing the best description. The Switching observer model represents priors and likelihood like the Bayesian observer, but instead of multiplying the two distributions, it switches between estimates chosen from the prior distribution and from sensory likelihood (hence its name), providing bimodal estimate distributions as a result. Finally, Sanborn & Chater (2016) provided an explanation for non-normative computations based on sampling processes. They addressed the dissociation between good reasoning in everyday life and poor performance in probabilistic experimental tasks suggesting that people are not 'ideal Bayesian reasoners' but they are Bayesian *samplers* working with a local sense of relative posterior probabilities and not with explicit Bayesian calculations. According to the authors, some phenomena observable in the human behavior, such as reasoning fallacies, stochasticity and autocorrelation, represent 'traces' of this sampling process.

2.2 Confirmation measures and their role in Bayesian epistemology

Works from psychology, decision making and behavioral economics highlighted and discussed the relevant role of posterior probability in decision making under uncertainty; on the other hand, when assessing the soundness of inductive arguments or the value of information, evidential impact is a crucial notion. In the psychological literature, the first notion was always given more attention than the latter, which, on the other hand, was a central concept for philosophers and epistemologists. In the epistemological literature, 'confirmation' defines the support that a certain piece of evidence provides towards a hypothesis. Thus, confirmation theory is "the study of the logic by which scientific hypotheses may be confirmed or disconfirmed (or supported or refuted) by evidence" (Hawthorne, 2011). In Eells & Fitelson 's (2002) words, "Measures of evidential support [...] are supposed to quantify the degree to which a piece of evidence E provides, intuitively speaking, "evidence for or against" or "support for or against" a hypothesis *H* – in an incremental as opposed to a final or absolute way" . Contemporary epistemology found in the Bayesian approach to probability a fitting candidate for uncertainty assessment, as it allows to mathematically *quantify* the degree of confirmation (or *degree of belief*) that a piece of evidence *e* provides for a hypothesis *H* in terms of mathematical probabilities (Hartmann & Sprenger, 2010). In this approach, indeed, confirmation is defined by the relationship between the conditional probability of a hypothesis *h* towards an evidence and prior probability of *h*. Estimation of confirmatory power is crucial in an inductive reasoning framework: after Carnap's formalization of 'confirmation' (Carnap, 1950), a large number of alternative measures of confirmation have been proposed and discussed in the epistemological literature (for a review of the different confirmation measures see Crupi, Tentori, & Gonzalez, 2007; Eells & Fitelson, 2002; Fitelson, 1999; Good, 1984; Tentori, Crupi, Bonini, & Osherson, 2007).

2.2.1. Psychology of inductive reasoning

One exception to this lack of interest for evidential impact in psychological research is the investigation of *categorical induction*. Categorical induction is the process allowing people to "arrive at a statement of their confidence that a conclusion category has a predicate after being told that one or more premise categories do" (Sloman & Lagnado, 2005). In this context, being able to estimate the strength or soundness of an argument is crucial. In the last twenty years, several explanatory accounts for inductive reasoning have been proposed. Heit (1998) provided a first Bayesian analysis of inductive reasoning showing how Bayes Theorem can be applied to evaluating an inductive argument; here, Bayesian analysis would rely on three assumptions to explain basic phenomena in inductive reasoning like similarity, typicality, and diversity effects as well as phenomena related to meaningful properties. This work addressed basic phenomena in inductive reasoning such as similarity, typicality and diversity effects. Heit's (1998) model was further discussed and empirically tested by Sanjana & Tenenbaum (2003), who described a Bayesian model of the probability of generalization which they tested on three data sets. They compared the relative ranking of the strengths of all arguments predicted by three theoretical models and provided by human participants and found that the Bayesian model showed a based rational foundation and quantitative advantage over the best similarity-models. A further formalization of induction as a form of Bayesian statistical inference over structured probabilistic models of the world was proposed by Tenenbaum, Griffiths, & Kemp (2006), who discussed inductive learning and reasoning in computational terms focusing on generalization, property induction and causal inference. To conclude, a review by Hayes et al. (2010) highlighted two relevant directions in research on induction. First, they point out the increasing attention that more articulated formal models are gaining, thanks to their explanatory power; second, they acknowledge that induction researchers are striving to examine the links between induction and other cognitive processes, thus turning their attention to a broader range of phenomena.

Overall, these works showed that probabilistic reasoning in everyday life is characterized by strategies that are normatively suboptimal but work well in a 'less than ideal' context. In the literature, different explanatory accounts of suboptimality have been proposed; some seem more compelling and provide more detailed explanations than others do.

2.2.2. Confirmation measures and surprisals

After almost exclusively focusing on the investigation of posterior probability for decades, in recent years psychologists have started to empirically test confirmation measures and compare them to probability judgments, moving the abovementioned theoretical investigation to a more empirical level. For example, Tentori, Crupi, Bonini, & Osherson (2007) described alternative measures of confirmation, discussed their normative appeal and compared their adequacy as description of confirmation judgments in a probabilistic context. After outlining some properties of these measures, the authors carried out an experimental test to compare their adequacy with respect to different formal criteria. For the experimental test, they set up an urn and ball task derived from early experiments in probability judgments to identify which of the confirmation measures corresponded most closely to evidential impact judgments provided by the participants. The results obtained showed that only one confirmation measure among those considered could satisfy necessary conditions on normative adequacy but warned about needing further research to understand its descriptive adequacy. Interestingly, the authors noted that people proved able to distinguish between posteriors and degrees of confirmation and suggested that evidential impact might be psychologically prior to probability. This hypothesis was then applied to conjunction fallacy (Crupi et al., 2008) and tested empirically on corpora-based predictions (Paperno, Marelli, Tentori, & Baroni, 2014) and inductive arguments (Tentori et al., 2016). We believe that also other works provided confirmation-based explanations for apparently suboptimal reasoning, labelling them in other ways. Itti & Baldi (2009) investigated how Bayesian surprise attracts attention by means of eye movements; to test their hypothesis, the authors recorded eye movements of experimental subjects while they watched a series of videoclips portraying dynamic natural scenes. While watching, participants were instructed to "follow the stimuli's main actors and actions, so that their gaze shifts reflected an active search for nonspecific information of subjective interest" (Itti & Baldi, 2009). By means of eye tracking data, the authors collected information on saliency, dynamic and surprise which allowed them to formalize a Bayesian definition of surprise "as the distance between the posterior and prior distributions of beliefs over models" (Itti & Baldi, 2009) based on two crucial concepts: *uncertainty* and *relativeness*. In light of these two features, a consistent definition of surprise must involve probabilities to cope with uncertainty, and prior and posterior distributions to reflect subjective expectations. The authors refer to "units of surprise" to quantify the variation between a prior probability and a conditional probability and for which, they believe, a quantitative measurement does not exist. However, in light of their definition of surprise we can consider it as a measure of *distance* and an intrinsically relative, comparative concept and relate it to confirmation. This idea of surprise was reprised by Prime & Shultz (2011) to explain the discrepancy between experimental failures and optimality in everyday life in probabilistic reasoning. However, in this other context, surprisal represents a mathematically defined concept taken from information theory that, according to the authors, captures people's intuitions about probability better than probability itself. In a behavioral experiment implemented on SurveyMonkey, the authors manipulated information (probability vs. frequency) and question format (probability vs. surprisal) to explore how different presentation modalities would affect participants' judgments. To do so, they elaborated four versions of Bayesian problems each giving and asking different information (probabilities vs frequencies). Participants were asked to read each problem and make a posterior judgment for each of them without making calculations but using their intuitive judgments instead. The results showed that people conform to Bayesian predictions by using both prior and likelihood information to update posteriors but deviate from Bayes rule when given frequency information and asked to provide answers in surprisals. In this case, they would use likelihood information but ignored prior probabilities. Overall, the effect of priors was much smaller than the effect of likelihoods when both were used; this tendency has already been described in the literature (Bar-Hillel, 1980). Overall, the authors noted that "people are more comfortable with judging their own surprise than with estimating probabilities even though "surprisals are not much used in psychological research, despite widespread psychological interest in manipulating and measuring surprise" (Prime & Shultz, 2011). According to the authors, then, surprisals are more intuitive and immediate than probabilities but at the same time they are scarcely used in psychological research; based on these two features, it is possible to draw a comparison between them and confirmation measures.

2.2.3. Probability and reasoning biases

Classic research in probabilistic reasoning, like the works from Kahneman & Tversky (Kahneman & Tversky, 1973, 1982; Tversky & Kahneman, 1974, 1981), argued that human reasoners' probability judgments systematically diverge from the normative benchmark and that people use suboptimal and biased strategies. In fact, a large part of the psychological literature on the topic has focused on describing and discussing a set of phenomena, so called *reasoning biases*, which were interpreted as suboptimal strategies of probabilistic reasoning. In an influential work, Tversky & Kahneman (1974) suggested that "people rely on a limited number of heuristic principles by which they reduce the complex tasks of assessing likelihoods and predicting values to simpler judgmental operations". They described four heuristics that are commonly employed to assess likelihoods and predict values (*representativeness*, *availability*, *adjustment* and *anchoring*), presented systematic biases to which these heuristics lead and discussed applied and theoretical implications. The authors ascribed heuristic judgments to limited cognitive resources and distinguished them from emotional and motivational factors affecting judgment. As already pointed out, the task environment of these early works contrasted with real life settings as it was often characterized by a deliberate dissociation between probability and confirmation.

During the years, several explanations for suboptimal reasoning have been proposed and discussed. Some of them relied on pragmatic issues and some others on cognitive heuristics; for the goals of the present research, I will focus on the information theoretic account proposed by Crupi and Tentori (Crupi et al., 2007; Crupi et al., 2008; Tentori, Chater and Crupi, 2016) which implies that evidential impact plays a crucial role in probabilistic inference.

2.3. Confirmation relations as possible explanations for reasoning biases

Evidential impact has been used to explain and contextualize some phenomena that classical experiments on probabilistic reasoning mentioned above tended to label as biases. One relevant example is *conjunction fallacy*, defined as the tendency "to regard a conjunctive event as more probable than one of its components, contrary to the conjunction rule of probability theory" (Kahneman and Tversky, 1973, 1982) caused by reliance on the representativeness heuristic. Some recent works presented the conjunction fallacy as an impact-driven probabilistic judgment conflicting with the normative benchmarks that classical works on probabilistic reasoning relied on. In a theoretical note aimed at discussing fallacious probability judgments entailed by this phenomenon, Crupi et al. (2008) provided an explanatory account based on the suggestion that these fallacious judgments are guided by sound assessment of confirmation relations (see Sides, Osherson, Bonini, & Viale, 2002). After describing the relationship between confirmation and probability in inductive logic and in the psychology of induction, the authors compared their confirmation-theoretic explanatory account to other relevant alternatives. From a theoretical point of view, the authors suggest that in reasoning biases like the conjunction fallacy, the notion of confirmation could overcome that of probability. Additionally, intuitive assessments of confirmation can be elicited directly and people can distinguish between probability and confirmation, as suggested in a more empirical work (Tentori et al., 2007). This confirmation-theoretic account was further investigated in a following study (Tentori et al., 2013) involving four experiments sharing the same basic procedure but with different elicitation procedures for confirmation and probability judgments, different experimental designs, classes of problems and content. The results of all four experiments supported the information-theoretic account and showed that "the perceived degree of confirmation [...] performed better than its perceived probability as a predictor of the occurrence and prevalence of the conjunction fallacy" (p.247). Moreover, when probability and confirmation were disentangled, the latter systematically prevailed as a determinant of the conjunction fallacy, indicating that inductive confirmation holds a crucial role in this phenomenon and further corroborating the confirmation-based explanation. Additionally, the authors pointed out that the results presented in this work could not be explained by alternative accounts like the ones discussed before. As mentioned before, some of these accounts seem more compelling and provide more detailed explanations than others do. In the next paragraph, I will present further evidence in line with the confirmation-based one proposed by Tentori et al. (2016). This same approach will also represent the theoretical basis for my experimental investigation of how confirmation and probability affect perceptual probabilistic inferences.

2.4. Combining evidential impact and probability judgments in psychological research

Classic works on probabilistic reasoning seemed to show that people did not perform well in experimental settings, whereas they had no problem with reasoning tasks in real life. Additionally, people seem to perform better in perceptuo-motor tasks than in cognitive decisions; for example, Jarvstad, Hahn, Warren, & Rushton (2014) addressed the issue suggesting that this dissociation might be due to several noise sources. By means of two experiments involving a perceptuo-motor task, the authors discovered that optimality was task dependent and thus statements about optimality should be more cautious. These two experiments involved an implicit as well as an explicit motor choice task and showed that participants' perceptuo-motor choices might deviate from optimality in two ways. First, they appear to prioritize speed over precision in the speed-accuracy trade-off; second, their precision criteria are less strict when aiming for larger targets.

Another explanatory account is the abovementioned confirmation-based one. In the last few years, some papers tried to compare directly confirmation and probability judgments to test which of the two is more accurate, coherent, and stable in time and therefore more likely to be the basis for probabilistic inference. I will focus on two relevant works, the first coming from a linguistics background and working with corpus-based data and the other one from a psychological background and involving real life arguments.

Paperno et al. (2014) demonstrated that confirmation affects human judgments on word co-occurrence likelihood. The aim of this work was to test whether corpus-based estimations are predictive of human intuitions on word probabilities. Three experiments involving different material but with a common experimental paradigm and procedure were ran. In all three of them, participants were asked to perform a forced choice between two candidate target words in the context of another word. Participants were asked to be as accurate as possible and to provide confidence ratings. Experimental stimuli were based on word co-occurrence data from a large text collection. Experiment 1 represented the first investigation of the effects of the confirmation between words on their perceived probability, comparing target pairs with equal conditional probabilities. Under these conditions, if participants' probability judgments were only driven by conditional probability, the choice rate for either target was expected to be at chance level. On the contrary, if confirmation relations affected probability judgments, participants' choice should be influenced by the evidential impact of context on target. Experiment 2 aimed at extending Experiment 1's investigation to a more general setting in which the two targets were not matched with respect to the conditional probability of the target in light of the context. To investigate whether and how confirmation values could affect likelihood judgment with targets whose posterior probabilities differed, Experiment 2 involved a random sample of items from the corpus where probability and confirmation varied freely, and used both variables as predictors in a regression analysis on participants' choices. Experiment 3 aimed at testing the effects described above under more controlled conditions.

In all three experiments, the authors found that confirmation consistently affected human judgments on word co-occurrence likelihood. The theoretical contributions of this work are manifold. First, it illustrated the usefulness of linguistic corpora as a source of probability and confirmation values implicit in language data; second, it contributed to the study of word association in linguistics; finally, it provided further evidence that speakers are sensitive to very subtle statistical patterns present in corpora.

The crucial role of confirmation relations in probability judgments was further explored with different experimental stimuli by Tentori, Chater and Crupi (2016) aimed at comparing the reliability of impact versus probability judgments in inductive inference. The goal of this study was to investigate whether impact judgments are more consistent and stable in time than probability estimates and to compare the reliability of impact versus probability judgments. To test this hypothesis, a paper and pencil experiment was carried out. Before the main experiment, a preliminary study was conducted to obtain response frequencies for real word arguments in order to derive objective probabilities and impact values to judge in the main experiment. Impact values were computed according to three different measures: probability ratio, likelihood ratio and relative distance. In a second phase, objective probabilities and impact values were used to generate 56 arguments by combining two complementary pieces of evidence with 28 hypotheses. Two pieces of evidence were used for each hypothesis to have an identical number of arguments with positive and negative impact. The proper task for each participant consisted in reading all the 56 arguments and judging the probability of the hypothesis in light of the evidence given and the impact of the evidence on the hypothesis. These judgments were repeated twice by every subject with an interval of around 7 days to obtain a time-consistency measure. The results showed that impact judgments were more consistent in time than probability judgments, independently from the measure considered. Impact judgments were also more correlated with objective probabilities computed from the preliminary survey than it was posterior probability. In general, these results suggest that impact judgments are more accurate and stable in time than posterior probability: they might be a more primitive type of judgment and might be the basis of probabilistic inference. These results align with older, theoretical works on confirmation measures discussed above showing that people are sensitive to the distinction between confirmation and probability.

2.5. Probabilities in perception and psychophysics

2.5.1. Statistical regularities in visual perception

The two abovementioned works by Paperno et al. (2014) and Tentori et al. (2016) investigated the effect of confirmation relations with verbal and linguistic material; as mentioned in the introduction, the goal of the present work is to understand whether and how confirmation relations affect probabilistic inferences even when perceptual (e.g. visual) material is at issue. Literature on visual perception often showed how it relies a lot on the identification and encoding of statistical features of the scene and how people can extract statistical measures over a variety of visual properties (Allan, Hannah, Crump, & Siegel, 2008; Ariely, 2001; Hannah, Crump, Allan, & Siegel, 2009; Yang, Tokita, & Ishiguchi, 2018; Zhao, Ngo, Mckendrick, & Turk-browne, 2011). Examples of this ability appear in works on ensemble statistics perception, exploring how the visual system naturally represents sets of similar items using summary statistics. In a review of works on summary statistical perception, Haberman & Whitney (2012) presented a variety of domains in which people are able to integrate or ensemble-code low level feature information, like position, size, orientation, or shadow. Ariely (2001) investigated the idea of set representation and explored whether the visual system creates a specific representation for a set of similar objects or it just encodes the sum of the representations of the individual items. To answer this question, two novel tasks were employed: member discrimination and mean identification. The first task measured knowledge about the sizes of individual spots in a set, whereas the second one measured sensitivity to the mean size of a set. In member-identification experiments, participants appeared to be unable to distinguish test spots that were in the set from those that were not and their performance was only marginally better than chance: in this task, observers did not seem to be able to make accurate judgments regarding parts of a set. In mean-discrimination experiments, on the other hand, participants reported much more accurate performances. Overall, then, the results of the mean discrimination experiments were characterized by a high level of accuracy, whereas member-identification tasks were not. Ariely (2001) provided an explanation for these results based on set representation, suggesting that the representation of a set is something more than a composition of its single parts and that people can extract statistical properties of a display. This set representation proposal was further explored by Corbett & Oriet (2011) in four behavioral experiments, involving the two tasks introduced by Ariely (2001) in combination with a RSVP (rapid serial visual presentation) paradigm. Overall, the results showed that explicit encoding of individual items is not necessary to build a mean representation of a set, further corroborating the statistical averaging hypothesis. Human participants, indeed, proved to be able to extract a representation of the mean size of a set of perceptual stimuli, but performed poorly when asked to determine whether a certain item was a member of this set. Taken together, these two results suggest that the visual system represents the overall statistical properties of sets of objects. Statistical representations, in turn, allow to perceive statistical regularities, which spontaneously guide and attract attention (Zhao, Al-Aidroos, & Turk-Browne, 2013) and reduce perceived numerosity by means of grouping mechanisms (Zhao & Yu, 2016).

The works from Ariely (2001) and Corbett and Oriet (2011) involved items characterized by one-dimensional features; multidimensional items, instead, are common in studies on visual statistical learning. Despite sharing their conceptual foundations (visual statistical processing), statistical summary perception and statistical learning are different processes: statistical summary perception involves the extraction of summary statistics over sets of objects; statistical learning, instead, involves the extraction of relationships among individual objects over repeated experience (Zhao et al., 2011). As I will describe in the next paragraph, these relationships have been conceptualized in different ways; I propose that confirmation relations could be one of them.

2.5.2. Parallels between contingency and confirmation relations

Confirmation and contingency relations seem to present several parallels and similarities from a formal and operational point of view; in light of this, an interesting hypothesis could be that the notion of confirmation relation can inform works on contingency judgments and help explain their results.

Contingency judgment tasks have been present in the literature since the eighties (Allan, 1980; Allan & Jenkins, 1983; Allan et al., 2008; Baker, Berbrier, & Vallee-tourangeau, 1989; Shanks, 1985); however, a crucial turning point is represented by the so-called *streamed-trial paradigm* which was first used by Crump, Hannah, Allan, & Hord (2007). They ran a behavioral experiment involving two types of judgments in a within-subject design: contingency and frequency ratings, related to sets of perceptual stimuli. This work established a novel procedure for measuring contingency judgments and replicated two central findings in contingency judgment literature: the dependence of contingency judgments on ΔP and on outcome density. ΔP represents a common measure for contingency and it is defined by the difference between the probability of an outcome given the cue and the probability of the same outcome in absence of the cue: $P(O|C) - P(O|\neg C)$. This equation is formally equivalent to confirmation measure n (Nozick, 1981): $n(e, H) = \Pr(e|H) - Pr(e|\neg H)$. In light of this equivalence, the parallel between confirmation and contingency measures already suggested represents a viable interpretation. Consequently, one could interpret the works on contingency judgments in terms of people assessing confirmation relations. Allan, Hannah, Crump, & Siegel (2008) further tested the streamed-trial paradigm in a series of four experiments, again investigating contingency assessment. In the four experiments, participants were asked to categorize the contingency between two items as either *weak* or *strong*, thus making a binary choice. Overall, these four experiments showed that parameters such as contingency sign, outcome density and payoffs affected decision criteria but not sensitivity to contingency and, more generally, that the streamed-trial task represented a viable methodology for contingency assessment. While discussing the results, the authors raise an interesting point: given that "psychophysics is the study of the relationship between physical events and our internal experience of these physical events" (Allan & Siegel, 2002), contingency judgments should represent a central research topic for psychophysics but this does not appear to be the case. Finally, Hannah et al. (2009) used the same streamed-trial procedure but modified it to investigate cue-interaction effect, which arises from pairing multiple cues with a common outcome: in this situation, people behave as if these cues interacted with each other instead of treating them independently. Once again, contingency was presented and computed in terms of ΔP measure. By means of three behavioral experiments involving geometric forms and more meaningful images implying some background knowledge, the authors were able to show that the streamed-trial procedure can be extended to the study of cue-interaction and that it can be used with different experimental materials. The works mentioned so far do not provide Bayesian explanatory accounts; McKenzie & Mikkelsen (2007), instead, proposed a Bayesian view of covariation assessment to explain two phenomena frequently found in covariation judgment: overestimating the weight of joint presence and underestimating the weight of joint absence and influence of prior beliefs on the variables' relationship judgment. Both phenomena represent departures from the normative model, but they are consistent with a Bayesian interpretation of the task. The authors presented the results of two behavioral experiments aimed at understanding whether the abovementioned phenomena appear in empirical judgments. These experiments revealed an inversion in the "cell A bias" (i.e. overestimating the weight of joint presence) when participants are led to believe that absence (instead of presence) is rare: cell D (joint absence) is then considered the most informative, in line with a Bayesian approach. While sensitivity to prior beliefs is a sign that people assess covariation in a "Bayes-like" way, sensitivity to rarity provides evidence for sensitivity to likelihoods and, consequently, one could infer, to confirmation relations. These results suggest that people take an inferential approach to covariation tasks instead of a descriptive one. That is, they try to assess the likelihood that there is a relationship between the variables involved. Because of this, they will tend to focus on instances of joint presence (cell A) and discard those of joint absence (Cell D). In a descriptive approach, instead, people would only focus on the four cell values without assuming any other external knowledge and considering all cells equally informative. In such a scenario, "cell A bias" would represent a deviation from the norm; however, in the inferential one preference for joint presence and the influence of prior probability generate from normative processes. Taken together, the results of both experiments provide empirical evidence for an inferential approach to covariation tasks providing further evidence for sound, despite not normative, inferences: "taking into account real-world conditions, combined with normative principles that make sense under these conditions, can help explain why people behave as they do" (McKenzie & Mikkelsen, 2007). Finally, Leshinskaya and Thompson-Schill (2018) found that participants in a statistical learning task involving perceptual stimuli are sensitive to not only the conditional probability between two events, but also the uniqueness of that relation. Once again, ΔP was used as measure of association; to test whether learners were sensitive to uniqueness in a visual statistical learning (VSL) task, the authors manipulated the *uniqueness* of a strongly predictive event pair in event sequences composed of animated events. To do so, they created low ΔP and high ΔP sequences. In the first case, they increased the conditional probability of the effect by having it follow two other events and itself more often than in the high ΔP sequence. Thus, the two conditions differed in terms of how uniquely the cause, rather than other events, predicted the effect. A first behavioral experiment carried out on Amazon Mechanical Turk involved a cover task in which participants were asked to identify the "common" versus "rare" version of each event type and showed that participants reported a weaker representation of the cause-effect relationship in the low ΔP condition; conversely, the high ΔP condition allowed participants to notice the predictive pattern. A second experiment was carried out to test the possibility that participants' performance could be affected by having only two alternative causes or many alternative causes for a frequent effect and showed that the amount and entity of alternative causes did not affect learning of the effect-cause relation. The key finding was that participants in a statistical learning task were sensitive to not only the conditional probability between two events, but also the uniqueness of that relation: uniqueness, then, could represent a crucial element also in associative learning and causal reasoning. Additionally, the authors found that the computation of such uniqueness happened incidentally and automatically, making it "the way we register the naturally occurring statistics of our observed world" (Leshinskaya and Thompson-Schill, 2018). Similarly, as proposed by Tentori et al. (2016), the perception of confirmation relations is also automatic and spontaneous.

Overall, these works showed parallels between the notion of contingency judgment and that of confirmation relations at a theoretical and empirical level. Thus, I propose that they represent analogous concepts in two different literatures and theoretical frameworks. That is, when performing contingency judgment tasks with streamed-trial or other procedures, people could base their answers on confirmation relations. It is also possible to draw a parallel between these studies and those reporting optimal probabilistic reasoning with apparently suboptimal computations, the common point being confirmation relations.

Chapter 3: The cognition-perception boundary: Bayesian brain hypothesis

Investigation of posterior probability and evidential impact has consequences and implications for a variety of topics which go beyond thinking and judgment specifically. In the last twenty years, probabilistic models of cognitive processes have gotten more and more popular and Bayesian probability theory has provided valuable contributions to different theoretical and empirical lines of research and scientific disciplines because it provides a theoretical framework for dealing with reasoning under uncertainty. A recent and widespread hypothesis is that of the so-called Bayesian brain (Knill & Pouget, 2004; Sanborn & Chater, 2016; Seriès & Seitz, 2013). According to this hypothesis, the brain represents sensory information in the form of probability distributions (Knill & Pouget, 2004) and "can be conceptualized as a probability machine that constantly makes predictions about the world and then updates them based on what it receives from the senses" (De Ridder, Vanneste, & Freeman, 2014). The Bayesian brain hypothesis has been applied to several low and high level processes like visual perception (Mamassian, Landy and Maloney, 2002), multisensory perception (Beierholm, Quartz and Shams, 2009), sensorimotor control (Körding & Wolpert, 2006), inductive learning (Tenenbaum et al., 2006). At a general level, it is well known in psychology, psychophysics and neuroscience that the nervous system of humans and animals developed to be sensitive to the statistical properties of the environment. Additionally, there is extensive evidence (e.g. see Beck et al., 2008; Chan, Niv, and Norman, 2016; Spratling, 2016; Vossel et al., 2015; and Wei and Stocker, 2012) in cognitive neuroscience and biology showing how probabilistic updating provides an almost perfect descriptive model for basic neural processes like vision and other kinds of perception by framing them in terms of probabilistic inferences over underlying probability distributions. Some computations are easier to frame in probabilistic terms than others: standard Bayesian models describe lower level processes more accurately than higher level ones. Indeed, plenty of such models have been proposed and tested in the last years for mechanisms like visual perception, attention and search, multisensory perception, perceptual decision making.

In light of the evidence coming from neuroscience and visual perception, there seems to be a discrepancy between normative benchmarks and people's reasoning performances; additionally, "People are Bayesians who fail to solve simple Bayesian word problems" (Sirota, Vallée-tourangeau, & Vallée-tourangeau, 2015). In other terms, people lack the ability to introspect about cognitive operations that are otherwise carried out in an optimal way in everyday life (Chater, Tenenbaum, & Yuille, 2006). This dissociation is reflected in two different approaches to the study of probabilistic reasoning involving different paradigms and apparently reaching different, if not opposite, conclusions. The first approach (Kahneman & Tversky, 1973) showed that people fail in simple Bayesian reasoning tasks, whereas a much more recent one (Griffiths & Tenenbaum, 2006; Sanborn & Chater, 2016; Tenenbaum et al., 2006) reported sound Bayesian reasoning in a variety of tasks. One work following this second approach (Griffiths & Tenenbaum, 2006) examined human cognition in more realistic context than laboratory studies and found that everyday cognitive judgments follow optimal statistical principles and align with the 'real' statistics in the world. This work discussed a behavioral experiment comparing ideal Bayesian analyses with the judgments of a large sample of participants, examining whether people's predictions were sensitive to the distributions of different quantities that arise in everyday contexts and whether they corresponded to optimal statistical inference in different settings. In the experiment, participants were asked to make predictions about five different phenomena (movie grosses, poem lengths, life spans, reigns of pharaohs, and lengths of marriages). Each prediction was based on one of five possible values, varied randomly between subjects. In each case, participants read several sentences establishing context and then were asked to give their predictions. People's judgments appeared to be close to the predictions coming from the Bayesian model across different settings, suggesting that people might be capable of considering prior distributions and update them in light of real world statistics.

3.2. Bayesian updating in perception and cognition

As discussed in the previous paragraph, plenty of studies focused on Bayesian reasoning with perceptual (e.g. visual) material as Bayesian inference is particularly suitable to model visual perception in terms of unconscious inference. Moreover, representing knowledge in terms of probability distributions is particularly suitable for low and high-level tasks involving uncertainty. Körding & Wolpert (2006) reviewed studies investigating the mechanisms involved in decision problems and action selection tasks, showed that human behavior aligns with Bayesian decision theory predictions and concluded that Bayesian decision theory represents a coherent framework for decisions involved in sensorimotor tasks. This framework can be applied to visual perception: Moreno-bote, Knill, & Pouget (2011) investigated whether visual percepts originate from Bayesian sampling, that is, sampling from probability distributions over image interpretations and showed that visual dominance in bistable perception behaves as a probability, supporting the idea of Bayesian sampling over a probability distribution. Bistable perception leads people to experience spontaneously perceptual alternation between two compelling interpretation of one single stimulus and represents one clear example of the interpretive nature of vision (Meng & Tong, 2004). As to more complex processes, Feldman (2014) reviewed Bayesian models of perceptual organization, first introducing the topic of Bayesian inference and then illustrating its application to perceptual organization problems. For our goals, the crucial part of this work is the definition of perception as unconscious inference, which explains why Bayesian inference has been proposed as a solution to this problem. Bayesian inference, indeed, deals with the notion of conditional probability (which we have defined above) in situations of *uncertainty*; thus, it represents an optimal candidate to model the central problem of perception which is to estimate physical world based on perceptual data. Rescorla (2015) provided a similar explanatory account and discussed the explanatory power and usefulness of mental representation in Bayesian perceptual models. If perception is framed as unconscious inference, expectations are a crucial element as they can be considered as prior beliefs in the inferential process (Seriès & Seitz, 2013). This idea is particularly relevant for the present work as it aims at drawing a parallel between inferential reasoning with verbal stimuli and visual perception. In their discussion of object perception as Bayesian inference, Kersten, Mamassian, & Yuille (2004) 'motivated' the Bayesian framework by once again framing perception as 'unconscious inference' in which one of the central goals is to make sense of ambiguity. In light of this conceptual framework, supporting psychophysical evidence and neural implications of the Bayesian approach to object perception are then discussed. Overall, the evidence provided in the paper suggests that the Bayesian framework is a fruitful scheme for studying object perception and presents several advantages, as it allows to explicitly model uncertainty, define ideal observers (and performance), and develop quantitative theories at the information processing level, and it applies to different areas such as language, speech, concepts and reasoning. A review by Seriès and Seitz (2013) discussed recent studies on motion perception in light of the perceptual Bayesian reasoning framework. The authors discussed how expectations can be described as probabilistic priors in a Bayesian updating framework and proposed two different types of effect on perception: expectations can modulate perceptual performance or they can alter the content of perception (i.e. the perceptual appearance of sensory inputs). In another review of Bayesian models of cognition, Chater, Oaksford, Hahn, & Heit (2010) describe perception as a 'prototypical' example of Bayesian inference because it aims at assigning probabilities to each possible interpretation of a percept, based on prior knowledge and sensory input. The Bayesian approach to perception is consistent with the idea that perception is "analysis by synthesis" (see also Yuille and Kersten, 2006); this means that the interpretation of perceptual data is the result of a combination of bottom up and top down processes. The idea of integration of priors and likelihood appears in the explanation of several different perceptual mechanisms as well as some other higher level ones. As Pouget and colleagues (2013) pointed out, "real-life problems are almost always far too complicated to allow for optimal behavior" and, as to visual perception, "natural images are both complex and objectively ambiguous" (Yuille & Kersten, 2006). In these cases, the brain might use heuristics or approximations: this is what makes behavior often suboptimal; other sources of suboptimality lie in the coding of sensory information and combination of sensory, perceptual and cognitive factors (Knill & Pouget, 2004). To deal with uncertainty, the brain models sensory data as conditional probability functions over a set of unknown variables. The idea of suboptimal inference appeared not only in decision making to explain cognitive biases, but also when describing perceptual decision making. According to Beck et al. (2012), suboptimal inference plays a relevant role in behavioral variability together with noise, especially when dealing with complex tasks. In these cases, the brain exploits computational shortcuts; therefore, most of behavioral variability comes from suboptimal inference due to these shortcuts. As to visual perception, it must be noted that the visual system treats perceptual sets with items varying along one dimension or more dimensions (i.e. conjunction of features). With stimuli varying along only one dimension (size, orientation) the overall statistics that can be computed are average, standard error and such, as described in the previous chapter. If stimuli represent the conjunction of two features, the computable overall statistics also include assessing relations between the two features across the set of objects. In Utochkin, Khvostov and Stakina (2018)'s words, "The variety of conjunctions as a function of their constituent feature statistics can be described in terms of inter-feature *correlation.* The correlation (or any other concordance measure) is an effective way to estimate how likely certain features in one dimension go with certain features in another dimension" (p.179). This idea of inter-feature correlation resonates a lot with the definition of visual statistical learning as "extraction of relationships among individual objects over repeated experience" outlined in the previous chapter; however, while the former seems to involve individual features, the latter explicitly mentions inter-object relations. Regardless of the level at which they work, both notions represent some kind of association between two items; given this assumption, a parallel can be drawn between inter-feature correlation and confirmation relations.

3.3. Criticism

Even though the Bayesian approach seems to provide a useful theoretical framework for the investigation of high and low level cognitive processes, some have raised critiques and observations about it as well as some doubts on its application to cognitive processes. In a review, Hahn (2014) distinguished three sets of criticism, by Jones and Love (2011), Elqayam and Evans (2011) and Bowers and Davis (2012a). These critiques focused on the unclear psychological implications of the model, its excessive flexibility with parameters and consequent unfalsifiability and its weak neuroscientific evidence. Jones & Love (2011) compared Bayesian approach to other psychological theories like behaviorism and evolutionary psychology, criticizing their use of optimality assumption. According to this work, explanatory status and theoretical contributions of Bayesian models of cognition can be easily brought back to already existing theories. The authors do not deny the importance and theoretical interest of Bayesian approach, but they warn against what they call 'Bayesian fundamentalism'. With this term, they referred to some research track whose primary goal is to "has been to demonstrate that human behavior in some task is rational with respect to a particular choice of Bayesian model" and which "strictly adheres to the tenet that human behavior can be explained through rational analysis […] without recourse to process, representation, resource limitations, or physiological or developmental data" (Jones & Love, 2011). Elaqayam & Evans (2011), instead, focused their criticism on *normativism,* "defined as the idea that human thinking reflects a normative system against which it should be measured and judged" and showed how this approach can lead to biased inference and does not represent a significant improvement to already existing computational level-analysis. Finally, Bowers & Davis (2012b) in their reply to Griffiths, Chater, Norris and Pouget (2012) criticized the large use of optimality claims by Bayesian researchers and raised three main arguments. First, they showed that the empirical evidence for Bayesian theories in psychology is weak; second, that this evidence is even weaker in neuroscience and finally they discussed the general Bayesian approach in cognitive science. According to Hahn (2014), these critiques are placed at the wrong level of generality: despite being motivated by specific models, the criticizing papers were misdirected as general critiques of a whole paradigm. Despite these criticisms, the Bayesian brain hypothesis still represents a very popular and widespread interpretation of lower and higher level cognitive processes.

As already mentioned, the two concepts of posterior probability and evidential impact are intrinsically related, and one implies the other: therefore, investigating whether and how they interact in perceptual probabilistic judgments would be useful and desirable and it could potentially cast light on some noteworthy phenomena, which goes beyond pure reasoning issues. At a more applied level, jointly understanding these two notions would be crucial for better understanding lower level processes (e.g. visual perception and search).

3.4. Experimental questions

As already mentioned, the hypothesis that the present thesis was aimed to test is that confirmation relations represent a viable explanatory account for apparently suboptimal probabilistic reasoning, both with verbal material (Tentori, Chater and Crupi, 2016) and perceptual stimuli. This experimental question is explicitly based on the confirmation-based model proposed by Tentori et al. (2016) but, at the same time, it can also inform the Bayesian brain Hypothesis. Indeed, if the brain "can be conceptualized as a probability machine that constantly makes predictions about the world and then updates them based on what it receives from the senses" (De Ridder et al., 2014) then instances of suboptimality could stem from humans using confirmation as a proxy for probability in their prediction-making process.

In light of the relevant role of impact on probability judgements, there are still open questions regarding the exploration of how people judge the posterior probability of a hypothesis $Pr(h|e)$ and the impact of new evidence on such hypothesis $Imp(h,e)$ in inductive inferences. For example, if we found that confirmation relations affect probability judgments even when perceptual material is at issue then it would mean that this relation also holds when no semantic content and no background knowledge is involved. The studies on contingency judgment cited in Chapter 2 already provide some evidence for confirmation-driven computations: in these works, in fact, people's judgments would often depend on ΔP, which is a confirmation measure itself. Most studies on contingency judgment do not adopt a completely abstract and blank experimental backstory, but they often involve some fictitious scenario (see Chapman & Robbins, 1990; Exp.2 in Hannah et al., 2009; Mandel & Lehman, 1998; and Vadillo, Miller & Matute, 2005). Because of this, it is possible to speculate that semantic background knowledge could still be involved. Additionally, to the best of my knowledge, none of these works has tried to empirically disentangle contingency and probability relations in order to explore whether one notion is more compelling than the other in driving judgments. To sum up, the theoretical background and empirical evidence discussed so far showed that confirmation relations affect probability judgments with verbal stimuli and word association judgments with linguistic corpora; the present dissertation aims at extending these results to a different context involving abstract, symbolic material. If we found significant effects even when perceptual stimuli are at issue, this would mean that the effect is transversal to at least three different experimental materials and is present even when background, semantic information is not available. With respect to the Bayesian models discussed above, this would mean that instead of turning to 'heuristic' shortcuts, people are indeed using Bayesian strategies (one could say that they 'are Bayesian'), but their focus is on confirmation relations instead of probability 'absolute' values.

3.4.1 Hypothesis

We operationalized the abovementioned experimental question in two hypotheses: a general one and a more in-depth one. In light of the first hypothesis, we expect participants to consider confirmed hypotheses more probable than corresponding (in terms of posterior probability) disconfirmed ones. If this were the case, it would mean that keeping all conditions constant, participants are influenced by confirmation relations despite being asked to perform a perceptual probabilistic task. As to the second hypothesis, we expect participants to choose the more likely option (i.e. the normatively correct one) more often when it is confirmed by the evidence provided than when it is disconfirmed. If this were the case, it would mean that in spite of being unaware of the conflict between impact-driven and normatively correct option, participants are affected by it when performing the task. Additionally, we also expect participants to be driven by confirmation relations when a normatively correct answer is absent. Finally, we expect different posterior probability levels to differently affect people's performance in the experimental task.

3.4.2 Strategy

To answer our experimental question and test the abovementioned hypotheses we ran four behavioral, computer-based experiments. They all shared the same underlying hypothesis and experimental paradigm but each of them involved different perceptual features to explore a potential generalizability of the effect.

Chapter 4: Experimental studies

Four behavioral computer-based experiments were conducted using the same methodology. All experimental stimuli were controlled using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997).

4.1 Experiment 1

4.1.1. Introduction

As discussed in the theoretical introduction, the few empirical works presenting a direct comparison between confirmation and probability estimates mostly involved verbal material and linguistic corpora. This kind of experimental material could raise issues related to the effect of previous information or semantic content on participants' performance; therefore, in order to minimize any background knowledge effect, we chose simple, perceptual features like color and shape. This first experiment represents a first exploration of whether and how evidential impact affects probability judgments when low-level features are at issue.

4.1.2. Method

Participants

Participants were 40 students of the University of Trento (25% men, mean age= 24, *SD*= 2.8). The study was approved by the ethics committee of the University of Trento, and informed consent was obtained for all participants. The sample size was calculated with G*Power (Version 3.1.5.1 Institut für Experimentelle Psychologie, Düsseldorf, Germany) assuming an effect size of 0.55, a α of 0.05, and a power of 0.8 ($1-\beta$). To determine the effect size, we chose to slightly overestimate the average effect size found in psychology (*d*= 0.4) in light of the high effect sizes (average $d = 1.65$) reported in a previous study by Tentori et al. (2016). The minimal sample size computed by this method was 28, therefore we aimed at recruiting at least 30 participants for experiment. However, in Experiment 1 we chose to recruit a much larger number as it represented a first exploration of the issue and were not sure that the task would be easily understandable by all participants. Having obtained satisfying performance from participants in Experiment 1, we kept a minimum of 30 participants in the following ones.

Material and Procedure

Experimental stimuli consisted of inductive arguments concerning 40 sets 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 white or black). The kinds of figures, their number, and their presentation time varied across experiments. Figures 3 to 6 in the appendix report all the 40 sets used in Experiment 1 as an example. As shown in Figure 2, features could be just displayed and/or verbally described, since for some of them, as pattern (Exp 3) line orientation or curvature (Exp 4), it was not possible to find an intuitive graphic representation that was uninformative about the levels of the other feature. By varying the probabilistic association between the two values of the features, we orthogonally manipulated posterior probability [three levels: .5-.5, .55-.45, and .6-.4] and impact [two levels: positive vs. negative]. A detailed description of experimental sets is provided in Figure 1. Sets were generated in order to have, for each possible combination of the two features, two arguments with the same posteriors and opposite impacts (i.e., equal in absolute strength but different in sign). This means that each value of the two features counted as evidence in 10 sets, and it was positively associated (in 5 sets, set "b" in Figure 1) and negatively associated (in 5 sets, set “c” in Figure 1) to the same level of the other feature, while keeping the posterior probability constant. As shown in Figure 2, in each trial, participants were first presented with a set of figures that had two features with two possible values each (i.e., white or black color and circular or triangular shape in Exp. 1, striped or dotted pattern and circular or triangular shape in Exp. 2, for examples, see 1a and 2a, 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). Participants were instructed to be as fast and accurate as possible and had 30 seconds to submit their answer. No immediate feedback on the accuracy of their responses was provided. Once the response was provided, the selected alternative appeared in the center of the screen for 1.5 seconds, and participants were prompted to press the spacebar to proceed with the next trial. Each set of figures was presented twice, for a total of 80 trials, with a 30 seconds break after every 20 trials. The experimental trials were preceded by four training trials. The presentation order of the trials was fully randomized. Participants were tested individually. At the end of the task, participants earned €0.15 for each correct answer. It is important to underline that the presentation modality is always above awareness level; this is relevant because, as demonstrated by Tapia, Breitmeyer, and Shooner (2010), among others, stimuli are processed at an individual-feature level at the nonconoscious level, but at a whole-object level while at the conscious level. By presenting all visual scenes for 3 to 4 seconds, we aimed at minimizing the probability that subjects would perceive only one of the two features composing each item.

a

	white	black		priors
triangle	12	12	24	0.50
circle	20	4	24	0.50
	32	16	**48**	
priors	0.667	0.333		

Pr(W|triangle)=	0.50	L(W, triangle)=	-0.33
Pr(B|triangle)=	0.50	L(B, triangle) =	0.33

b

	white	black		priors
triangle	12	10	22	0.48
circle	22	2	24	0.52
	34	12	**46**	
priors	0.739	0.261		

Pr(W|triangle)=	0.55	L(W, triangle)=	-0.40
Pr(B|triangle)=	0.45	L(B, triangle) =	0.40

c

	black	white		priors
triangle	12	10	22	0.48
circle	4	20	24	0.52
	16	30	**46**	
priors	0.348	0.652		

Pr(B|triangle)=	0.55	L(B, triangle) =	0.38
Pr(W|triangle)	0.45	L(W, triangle)=	-0.38

Figure 1: examples of experimental sets. In each contingency table, rows represent the two possible evidences and columns represent the two hypotheses. Pr: *posterior probability;* L: *evidential impact, according to L confirmation measure (Kemeny & Oppenheim, 1952). In set* a, *the two hypotheses (white and black) have the same posterior probability in light of the evidence "triangle"; thus, there is no normatively correct answer. However, evidence "triangle" provides positive support for hypothesis "black" and negative support for hypothesis "white". In set* b, *the posterior probability of hypothesis "white" in light of evidence "triangle" is higher than that of hypothesis "black". However, evidence "triangle" provides positive support for hypothesis "black" and negative support for hypothesis "white". In this set, then, the evidence provided disconfirms the more likely alternative. In set* c, *the posterior probability of hypothesis "black" in light of evidence "triangle" is higher than that of hypothesis white, and evidence "triangle" provides positive support for hypothesis "black" and negative support for hypothesis "white". In this set, the evidence provided also confirms the more likely alternative. For a complete list of the experimental sets involved in Experiment 1, see Appendix. Experiment 2 involved the same sets but white was replaced with "light gray" and black with "dark gray". In Experiments 3 and 4, we halved the number of figures in each set but probability and impact values were not changed.*

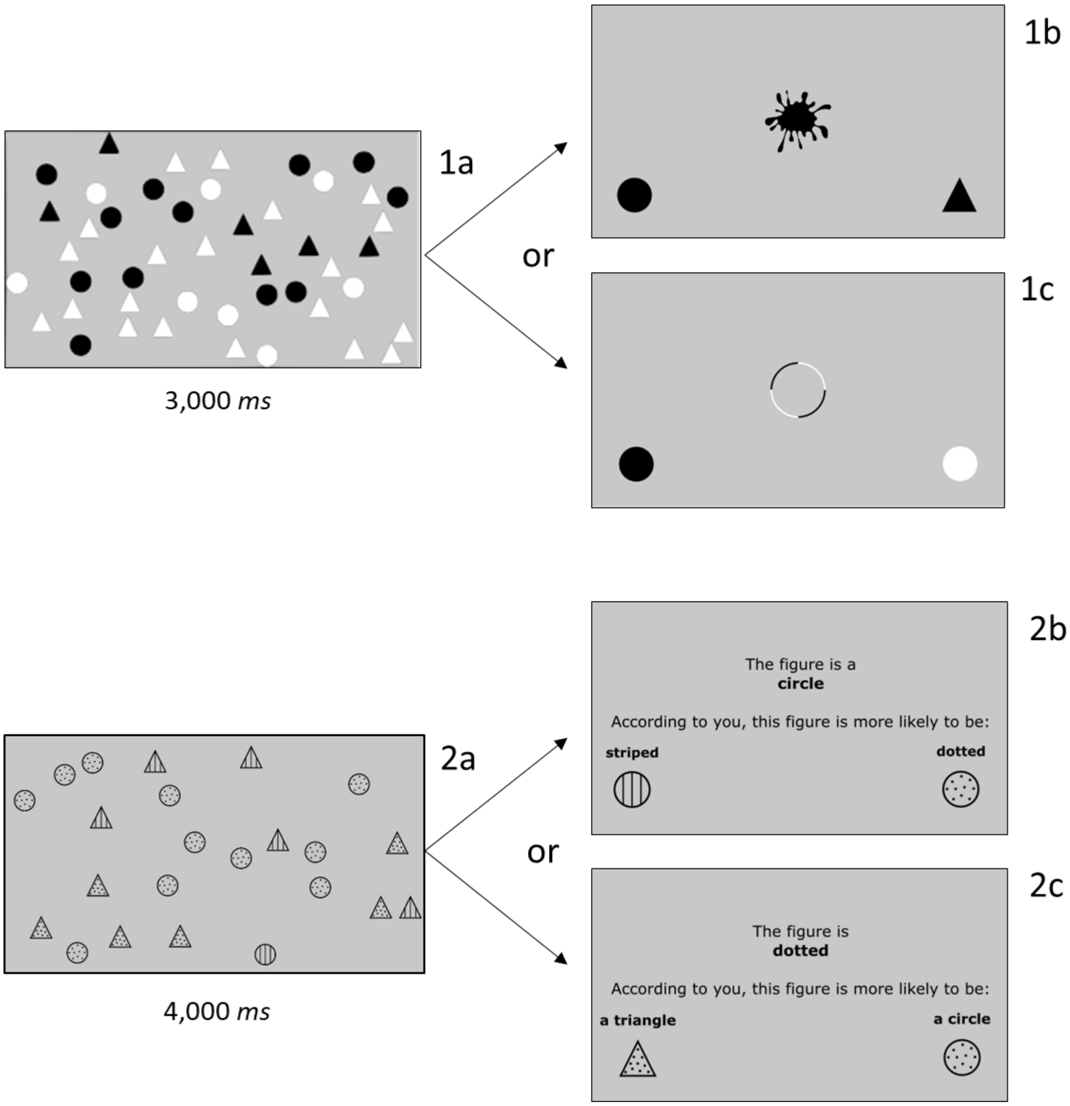

Figure 2: visual description of the experimental task. Participants were seated in a dimly illuminated room 60 cm from the monitor (1920x1080 resolution, 100 Hz). The generation and presentation of the stimuli was controlled by using Matlab and Psychtoolbox-3 (Brainard, 1997; Pelli, 1997). All the items of a given set were inscribed in a 2.6° by 2.6° square. Rectangles 1a to 1c and 2a to 2c exemplify experiments 1 and 3 respectively. The structure of experiment 2 was analogous to that of experiment 1 and experiment 4 was analogous to experiment 3. Features could be just displayed and/or verbally described, since for some of them, as pattern (exp 3) line orientation or curvature (exp 4), it was not possible to find an intuitive graphic representation that was uninformative about the levels of the other feature.

4.1.3. Data analysis

The same analytical strategy was employed for the four experiments. First, for each participant, we computed the proportion of choices for the confirmed alternative out of the total number of trials. In order to investigate if this proportion was higher than chance, we performed a one-sample *t*-test (against .5). Second, only for trials in which posterior probability differed from .5 (that is, trials in which there is a normatively correct response), we compared the proportions of errors for confirmed versus disconfirmed alternatives. More specifically, for each participant, we computed the proportions of choices for the less likely alternative out of the total amount of trials in which the evidence confirmed the more likely hypothesis and out of the total amount of trials in which the evidence disconfirmed the more likely hypothesis. These two proportions were then compared using a paired sample *t*-test. Third, a GEE regression analysis was used to model the effect of posterior probability and type of evidence (i.e., the feature provided as evidence) on choices for the confirmed alternative. Both factors were included in the analysis as categorical independent variables. Fourth, a GEE regression analysis was carried out to ascertain the effect of impact direction, posterior probability, and type of evidence on errors. All three factors were included in the analysis as categorical independent variables. When needed, we compared the number of choices for the confirmed alternative (or of errors) within specific sub-classes of stimuli using post-hoc pairwise comparisons with Bonferroni correction of the p values. These four analyses were repeated also for consistent trials that is only for trials in which the participant provided the same answer to the same two sets of figures. Data analysis was conducted using SPSS statistical analysis software (version 21).

4.1.4. Results

Table 1. Mean proportions of choice for the confirmed alternative in Experiment 1

Posterior Probability	Evidence	All trials Prop.	Consistent trials Prop.	Consistent trials %
0.5-0.5	Shape			
	Triangle	.52	.56	65
	Circle	.48	.44	44
	Color			
	Black	.57	.69	44
	White	.69	.81	66
0.55-0.45	Shape			
	Triangle	.52	.58	64
	Circle	.41	.28	45
	Color			
	Black	.68	.72	69
	White	.74	.89	57
0.6-0.4	Shape			
	Triangle	.48	.49	67
	Circle	.32	.15	50
	Color			
	Black	.74	.83	68
	White	.71	.72	77
Overall		.57	.60	61

"*Prop.*" stands for the mean proportion of choices for the confirmed alternative across participants in all trials (column "*All trials*") or consistent trials only (column "*Consistent trials*"). These proportions have been computed by averaging the proportion of choices for the confirmed option provided by each participant in each subclass of stimuli. "*%*" stands for the percentage of consistent trials out of the total amounts of trials in each subclass of stimuli

Table 2. Mean proportions of errors in Experiment 1

Posterior Probability	Evidence	All trials		Consistent trials			
		Disconfirmed	Confirmed	Disconfirmed		Confirmed	
		Prop.	Prop.	Prop.	%	Prop.	%
0.55-0.45	Shape						
	Triangle	.38	.33	.34	61	.28	66
	Circle	.37	.56	.20	42	.62	47
	Color						
	Black	.61	.24	.61	66	.17	72
	White	.71	.23	.90	54	.8	61
0.6-0.4	Shape						
	Triangle	.27	.31	.20	70	.19	65
	Circle	.28	.64	.14	59	.85	41
	Color						
	Black	.67	.19	.71	69	.4	67
	White	.57	.14	.55	71	.7	84
Overall		.48	.33	.46	62	.27	63

"*Prop.*" stands for the mean proportion of errors across participants in trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence, when all trials (column "*All trials*") or only consistent trials (column "*Consistent trials*") are considered. These proportions have been computed by averaging the proportion of errors made by each participant in each subclass of stimuli. "*%*" stands for the percentage of consistent trials out of the total amounts of trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence in each subclass of stimuli.

Table 1 and 2 report the percentage of choices for the confirmed alternative (Table1 1) and the percentage of errors (Table 2), in each of the sub-classes of stimuli. The one sample t-test revealed that the proportion of choices for the confirmed alternative out of the total number of trials ($M = .57$, SD = .14) was significantly higher than chance level [$t(39) = 3.34$, $p = .002$]. Coherently with this result, the paired sample t-test revealed that the proportion of errors out of the total number of trials was greater when impact and posterior probability were dissociated, that is when the more likely hypothesis was disconfirmed rather than confirmed by the evidence [$M = .48$, SD = .15, and $M = .33$, SD = .15, respectively, $t(39) = 3.54$, $p = .001$]. The results of the GEE regression analysis on choices for the confirmed alternative showed that the type of evidence was a significant predictor [$\chi^2(1) =$

49.49, $p < .001$] while posterior probability was not [$\chi^2(2) = 2.10$, $p = .350$]. More specifically, across various posteriors levels, participants showed an increase of choices for the confirmed alternative when the type of evidence provided was color rather than shape ($M = .69$, SE = .027, and $M = .46$, SE = .027; post-hoc test: $p < .001$). The GEE also showed a significant interaction between type of evidence and posterior probability [$\chi^2(2) = 16.39$, $p < .001$]. As shown in Table 2, the frequency of choices for the confirmed alternative tended to increase consistently with the increase of the posterior probability of the more likely hypothesis. However, when the evidence concerned the shape of the figure such tendency was significantly reversed. Participants showed a lower number of choices for the confirmed alternative with .60-.40 posterior ($M = .40$, SE = .025) rather than with .55-.45 ($M = .47$, SE = .03) and .50-.50 ($M = .50$, SE = .036) posteriors (post-hoc tests: $p =.035$, and $p = .002$, respectively).

The GEE regression analysis on errors showed that impact direction and posterior probability were both significant factors [$\chi^2(1) = 12.21$, $p < .001$, and $\chi^2(1) = 9.22$, $p = .002$, respectively], while type of evidence was not ($p = .837$). As expected, participants showed a decrease of errors when the evidence confirmed rather than disconfirmed the more likely hypothesis ($M = .32$, SE = .025, and $M = .48$, SE = .026, respectively; post-hoc test: $p < .001$), and with .60-.40 posteriors rather than with .55-.45 posteriors ($M = .37$, SE = .017, and $M = .42$, SE =.013, respectively; post-hoc test: $p < .001$). Significant interaction effects were obtained between impact direction and type of evidence [$\chi^2(1) = 62.43$, $p < .001$], and among all three factors [$\chi^2(1) = 5.89$, $p = .015$]. Across posterior probability levels, when the type of evidence was shape, the frequency of errors was slightly greater when the evidence confirmed rather than disconfirmed the more likely hypothesis ($M = .46$, SE = .031, and $M = .33$, SE = .031, respectively; post-hoc test: $p = .046$). Such a difference, however, was significant only with .6-.4 posteriors ($M = .48$, SE = .033, and $M = .28$, SE = .034, respectively; post-hoc test: $p = .001$).

Table 3. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (all trials) in Experiment 1

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,538	,1395	,264	,811	,000
posterior= 0.6-0.4[1]	,440	,1573	,131	,748	,005
posterior= 0.55-0.45[1]	,362	,1445	,079	,646	,012
type_of_evidence[2]	-,525	,1813	-,880	-,170	,004

[1] reference category: 0.5-0.5
[2]reference category: color

Table 4. Regression coefficients for the GEE regression on the proportion of errors (all trials) in Experiment 1

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,647	,1564	,340	,953	,000
posterior [3]	-,162	,1747	-,505	,180	,353
type_of_evidence[2]	-1,144	,2213	-1,578	-,710	,000
impact_direction[4]	-1,830	,2945	-2,408	-1,253	,000

[2]reference category: color
[3]reference category: 0.55-0.45
[4]reference category: negative

A very similar pattern of results was obtained when only consistent trials were considered. The proportion of choices for the confirmed alternative ($M = .60$, SD = .21) was significantly higher than chance level [$t(39) = 3.02$, $p = .004$], and the proportion of errors was greater when the more likely hypothesis was disconfirmed by the evidence than when it was confirmed [$M = .46$, SD = .26, and $M = .27$, SD = .24, respectively, $t(39)= 2.93$, $p = .006$]. Type of evidence resulted a significant factor [$\chi^2(1) = 50.34$, p < .001] in predicting choices for the confirmed alternative, while posterior probability did not [$\chi^2(2) = 5.08$, $p = .079$]. Again, the interaction effect between the two factors was significant [$\chi^2(2) = 8.67$, $p =$

.013]. The GEE analysis on errors reported significant effects of posterior probability [$\chi^2(1) = 6.53$, $p = .011$] and impact direction [$\chi^2(1) = 10.45$, $p < .001$], but not of type of evidence, (p = .813). The only significant interaction effect proved to be the one between impact direction and type of evidence [$\chi^2(1) = 67.81$, $p < .001$].

Table 5. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (consistent trials) in Experiment 1

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	1,039	,3065	,438	1,640	,001
posterior= 0.6-0.4[1]	,419	,2904	-,150	,988	,149
posterior= 0.55-0.45[1]	,565	,2882	,000	1,130	,050
type_of_evidence[2]	-1,016	,3672	-1,736	-,296	,006

[1] reference category: 0.5-0.5
[2]reference category: color

Table 6. Regression coefficients *for the GEE regression on the proportion of errors (consistent trials) in Experiment 1*

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	1,155	,2718	,622	1,688	,000
posterior[3]	-,448	,2559	-,950	,053	,080
type_of_evidence[2]	-2,175	,4090	-2,976	-1,373	,000
impact_direction[4]	-3,321	,5348	-4,370	-2,273	,000

[2]reference category: color
[3]reference category: 0.55-0.45
[4]reference category: negative

4.1.5. Discussion

In line with our hypothesis, participants systematically chose the confirmed alternative over the equally probable, but disconfirmed one, and chose the normatively incorrect (i.e. less likely) alternative more often when it was confirmed (vs. disconfirmed) by the evidence provided. However, the type of evidence being tested appeared to affect choice patterns significantly and interacted with the relative posterior probabilities and impact direction. That is, participants' choice patterns appeared to be affected by impact relations only when the more salient feature (in this case, color) was provided as evidence.

These results provided a first empirical evidence of the effect of confirmation relations on probability judgment with perceptual stimuli, but also highlighted a significant influence of the experimental material itself on choice patterns. In fact, in line with empirical findings that indicate a privileged role of color in visual search (Wolfe & Horowitz, 2004, 2017) as well as in modulating attentional capture (Adamo, Wozny, Pratt, & Ferber, 2010), the obtained results showed that color was a more compelling evidence than shape in determining participants' choices.

4.2. Experiment 2

4.2.1. Introduction

Having found that the task was easily understandable by all participants, we reduced the sample size. Additionally, having noticed that color represented a more compelling evidence than shape, we still used it as one of the two crucial features of each item, but in an attempt to make it less salient than before we used two shades of gray (dark and light) instead of black and white. Lastly, we added a thick, bright green rim to each item to make its shape more salient. The purpose was to force participants to also pay attention to the shape of each item.

4.2.2. Method

Participants

Participants were 30 students of the University of Trento (33% men, mean age= 22, *SD* = 2.8). The study was approved by the ethics committee of the University of Trento, and informed consent was obtained for all participants.

Material and Procedure

The experimental procedure was analogous to that described in Experiment 1, with the changes mentioned in the introduction: each item presented in the visual sets represented the conjunction of shape (triangle/circle) and color (dark/light gray).

4.2.3. Data analysis

The analytical strategy was analogous to that described in Experiment 1.

4.2.4. Results

Table 7. Mean proportions of choice for the confirmed alternative in Experiment 2

Posterior Probability		All trials	Consistent trials	
	Evidence	Prop.	Prop.	%
0.5-0.5	Shape			
	Triangle	.47	.46	58
	Circle	.44	.38	60
	Color			
	Dark	.71	.79	62
	Light	.70	.78	72
0.55-0.45	Shape			
	Triangle	.45	.44	58
	Circle	.39	.31	59
	Color			
	Dark	.75	.87	65
	Light	.67	.78	65
0.6-0.4	Shape			
	Triangle	.38	.31	68
	Circle	.37	.30	65
	Color			
	Dark	.74	.86	68
	Light	.71	.78	69
Overall		.56	.60	64

"*Prop.*" stands for the mean proportion of choices for the confirmed alternative across participants in all trials (column "*All* *trials*") or consistent trials only (column "*Consistent trials*"). These proportions have been computed by averaging the proportion of choices for the confirmed option provided by each participant in each subclass of stimuli. "%" stands for the percentage of consistent trials out of the total amounts of trials in each subclass of stimuli.

Table 8. Mean proportions of errors in Experiment 2

Posterior Probability	Evidence	All trials: Disconfirmed Prop.	All trials: Confirmed Prop.	Consistent trials: Disconfirmed Prop.	Consistent trials: Disconfirmed %	Consistent trials: Confirmed Prop.	Consistent trials: Confirmed %
0.55-0.45	Shape						
	Triangle	.37	.47	.32	63	.43	53
	Circle	.27	.50	.15	70	.48	50
	Color						
	Dark	.73	.23	.90	60	.15	70
	Light	.69	.35	.79	65	.27	67
0.6-0.4	Shape						
	Triangle	.19	.43	.7	75	.38	60
	Circle	.17	.43	.3	73	.36	57
	Color						
	Dark	.70	.22	.81	70	.11	65
	Light	.68	.24	.74	67	.15	72
Overall		.47	.36	.48	68	.29	62

"*Prop.*" stands for the mean proportion of errors across participants in trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence, when all trials (column "*All trials*") or only consistent trials (column "*Consistent trials*") are considered. These proportions have been computed by averaging the proportion of errors made by each participant in each subclass of stimuli. "%" stands for the percentage of consistent trials out of the total amounts of trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence in each subclass of stimuli.

Table 7 reports the percentages of choice for the confirmed alternative within all trial categories out of both the total amount of trials and consistent trials only. When all trials were considered, the one-sample t-test showed that the proportion of confirmed choices was significantly higher than chance level [$M = .56$, SD = .16; $t(29) = 2.25$, $p = .032$]. Coherently, the paired samples t-test on the proportion of errors showed that this proportion was higher when the more probable hypothesis was disconfirmed by the evidence provided than in the other condition confirmed [$M = .47$, SD = .15, and $M = .36$, SD = .20], but this difference was not statistically significant, $t(29) = 1.90$, $p = .067$. The GEE regression analysis on choices for the confirmed alternatives did not report a significant effect of posterior probability [$\chi^2(2) = 1.58$, $p = .454$] but a strong effect of the type of evidence provided [$\chi^2(1) = 55.68$, $p < .001$]. In fact, confirmed alternatives were chosen significantly more often when the evidence provided was color rather than shape ($M = .71$, SE = .038, and $M = .42$, SD = .028, respectively; post-hoc test: $p < .001$). The interaction between the two variables affected significantly the proportion of confirmed choices, $\chi^2(2) = 9.05$, $p = .01$]. The GEE analysis on the proportion of errors reported a significant effect of posterior probability [$\chi^2(1) = 12.74$, $p < .001$] and type of evidence [$\chi^2(2) = 27.21$, $p < .001$]. In fact, errors were less frequent with .6-.4 (vs .55-.45) posteriors and when shape was the evidence provided, rather than color. The direction of impact did not affect the proportion of errors significantly [$p = .109$] but its interaction with type of evidence did, [$\chi^2(1) = 69.21$, $p < .001$]. Consequently, when the evidence provided was shape, impact direction did not affect the proportion of errors, whereas when it was color, errors were more frequent when the evidence disconfirmed the more likely hypothesis, rather than confirm it ($M = .70$, SE = .042, and $M = .26$, SE = .042, respectively; post-hoc test: $p < .001$). Other two-way interactions effects were not found [all ps $> .05$] but the three way interaction between impact direction, posterior probability and type of evidence was significant, $\chi^2(1) = 4.52$, $p = .034$. Across posterior probability levels, when the type of evidence was shape, the frequency of errors was greater when the evidence confirmed rather than disconfirmed the more likely hypothesis ($M = .46$, SE = .034, and $M = .24$, SE = .029, respectively; post-hoc test: $p = .001$). Such a difference, however, was significant only with .6-.4 posteriors (confirming evidence: $M = .43$, SE = .040; disconfirming evidence: $M = .18$, SE = .023, post-hoc test: $p < .001$).

Table 9. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (all trials) in Experiment 2

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,867	,1929	,489	1,245	,000
posterior=0.6/0.4[1]	,102	,1312	-,155	,359	,436
posterior=0.55/0.45[1]	,030	,1309	-,226	,287	,818
type_of_evidence	-1,034	,2053	-1,437	-,632	,000

[1] reference category: .50-.50
[2]reference category: color

Table 10. Regression coefficients for the GEE regression on the proportion of errors (all trials) in Experiment 2

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,908	,2077	,500	1,315	,000
posterior[1]	-,138	,1727	-,477	,200	,423
type_of_evidence[2]	-1,677	,2753	-2,216	-1,137	,000
impact_direction[3]	-1,795	,4055	-2,590	-1,000	,000

[1] reference category: .50-.50
[2]reference category: color
[3]reference category: negative

When only consistent trials were considered, the proportion of choices for the confirmed alternative was still significantly higher than chance level $M = .59$, SD = .21; $t(29) = 2.25$, $p = .032$]. Unlike when considering all trials, in this case the proportion of errors when the more probable hypothesis was disconfirmed by the evidence provided was significantly higher than in the other condition [$M = .48$, SD = .19, and $M = .29$, SD = .29, $t(29) = 2.34$, $p = .026$]. The GEE analysis on the proportion of choices for the confirmed alternative reported the same results as with all trials: posterior probability did not affect choice patterns, $p = .482$, but type of evidence did, $\chi^2(1) = 44.50$, $p < .001$. Indeed, across posterior probability levels, participants chose the confirmed alternative more frequently when the evidence provided was color ($M = .82$, SE = .051) rather than shape ($M = .36$, SE

= .045; post-hoc test: $p < .001$). The two factors did not interact significantly, $p = .077$. A GEE analysis on the effect of posterior probability, type of evidence and impact direction on the proportion of errors reported main effects for posterior probability and type of evidence [$\chi^2(1) = 12.92$, $p < .001$ and $\chi^2(1) = 27.31$, $p <. 001$, respectively] but not for impact direction, $p = .144$. As with all trials, the two- way interaction between impact direction and type of evidence was the only statistically significant one [$\chi^2(1) = 45.91$, $p = .001$]. When the evidence provided was shape, impact direction did not affect the proportion of errors, whereas when it was color, errors were more frequent when the evidence disconfirmed the more likely hypothesis, rather than confirm it ($M = .81$, SE = .061, and $M = .15$, SE = .051, respectively; post-hoc test: $p < .001$). Other two-way interactions effects were not found [all ps $> .05$], but the three way interaction between impact direction, posterior probability and type of evidence was marginally significant, $\chi^2(1) = 4.13$, $p = .042$.

Table 11. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (consistent trials) in Experiment 2

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	1,451	,3540	,757	2,145	,000
posterior=0.6/0.4[1]	,130	,2312	-,323	,583	,575
posterior=0.55/0.45[1]	,077	,2611	-,435	,588	,769
type_of_evidence[2]	-1,739	,3538	-2,432	-1,045	,000

[1] reference category: .50-.50
[2]reference category: color

Table 12. Regression coefficients for the GEE regression on the proportion of errors (consistent trials) in Experiment 2

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	1,658	,4650	,747	2,570	,000
posterior=0.6/0.4[1]	-,460	,3462	-1,138	,219	,184
type_of_evidence[2]	-2,895	,5300	-3,934	-1,856	,000
Impact direction[3]	-3,075	,8107	-4,664	-1,486	,000

[1] reference category: .50-.50

[2]reference category: color
[3]reference category: negative

4.2.5. Discussion

As in the previous experiment, choice patterns generally align with our experimental hypotheses but also reveal a strong effect of the feature being provided as evidence. Indeed, once again, the type of evidence affected the proportion of choices for the confirmed alternative as well as the proportion of errors, and it interacted significantly with the relative posterior probability for the two hypotheses and with impact direction. Overall, these results seem to support our hypothesis but at the same time are affected by the type of evidence provided even more than in the previous experiment: instead of helping participants focus not only on the color of all items but on their shape as well, these new features seemed to make color even more salient. Moreover, error patterns for those sets where shape was the evidence almost seemed to suggest that participants were not following any particular strategy when answering to those trials.

4.3. Experiment 3

4.3.1. Introduction

The results of Experiments 1 and 2 suggest that individuals' probability judgments are influenced by evidential impact also when perceptual stimuli are at issue. In order to avoid the asymmetry between the two perceptual features observed in both of them, we tried to balance the saliency of type of evidence by replacing color with pattern (i.e., lines vs. dots). Moreover, to simplify the perceptual task, we halved the number of figures included in each set and incremented the presentation time of each trial.

4.3.2. Method

Participants

Participants were 32 students of the University of Trento (31% men, mean age= 20.4, *SD* = 2.2). The study was approved by the ethics committee of the University of Trento, and informed consent was obtained for all participants.

Material and Procedure

The procedure was analogous to that described in Experiment 1, with the changes mentioned in the Introduction; each item presented in the visual sets represented the conjunction of shape (triangle/circle) and pattern (lines vs dots).

4.3.3. Data analysis

The analytical strategy was analogous to that described in Experiment 1.

4.3.4. Results

Table 13. Mean proportions of choice for the confirmed alternative in Experiment 3

Posterior Probability		All trials	Consistent trials	
	Evidence	Prop.	Prop.	%
0.5-0.5	Shape			
	Triangle	.50	.52	56
	Circle	.42	.37	56
	Pattern			
	Lines	.66	.76	61
	Dots	.69	.86	59
0.55-0.45	Shape			
	Triangle	.48	.49	61
	Circle	.54	.59	53
	Pattern			
	Lines	.61	.78	50
	Dots	.70	.53	69

0.6-0.4	Shape			
	Triangle	.49	.47	70
	Circle	.50	.48	69
	Pattern			
	Lines	.68	.79	66
	Dots	.66	.75	66
Overall		.58	.61	62

"*Prop.*" stands for the mean proportion of choices for the confirmed alternative across participants in all trials (column "*All trials*") or consistent trials only (column "*Consistent trials*"). These proportions have been computed by averaging the proportion of choices for the confirmed option provided by each participant in each subclass of stimuli. "*%*" stands for the percentage of consistent trials out of the total amounts of trials in each subclass of stimuli

Table 14. Mean proportions of errors in Experiment 3

Posterior Probability	Evidence	All trials		Consistent trials			
		Disconfirmed	Confirmed	Disconfirmed		Confirmed	
		Prop.	Prop.	Prop.	%	Prop.	%
0.55-0.45	Shape						
	Triangle	.35	.38	.26	61	.26	61
	Circle	.44	.37	.39	48	.25	58
	Pattern						
	Lines	.59	.34	.69	45	.16	55
	Dots	.73	.29	.83	70	.17	67
0.6-0.4	Shape						
	Triangle	.31	.34	.24	66	.30	73
	Circle	.28	.28	.21	78	.10	59
	Pattern						
	Lines	.62	.24	.72	59	.14	73
	Dots	.62	.26	.69	58	.17	73
Overall		.49	.31	.49	61	.20	65

"*Prop.*" stands for the mean proportion of errors across participants in trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence, when all trials (column "*All trials*") or only consistent trials (column "*Consistent trials*") are considered. These proportions have been computed by averaging the proportion of errors made by each participant in each subclass of stimuli. "*%*" stands for the percentage of consistent trials out of the total amounts of trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence in each subclass of stimuli.

When all trials were considered, the one-sample t-test showed that the proportion of choices for the confirmed alternative ($M = .58$, SD $= .16$) was significantly higher than chance level [$t(31) = 2.69$, $p = .011$]. The proportion of errors when the evidence disconfirmed the most likely hypothesis was significantly greater than when the evidence confirmed it [$M = .49$, SD = .18, and $M = .31$, SD = .18, respectively, $t(31) = 3.28$, $p = .003$]. The GEE regression analysis on choices for the confirmed alternative showed a strong effect of type of evidence [$\chi^2(1) = 50.29$, $p < .001$], but no significant effect of posterior probability [$\chi^2(2) = .44$, $p = .802$], and no interaction between type of evidence and probability [$\chi^2(2) = 3.18$, $p = .204$]. Across posterior probability levels, participants chose the confirmed alternative more frequently when the evidence provided was pattern ($M = .67$, SE = .027) rather than shape ($M = .49$, SE = .020; post-hoc test: $p <. 001$). The GEE regression analysis on errors showed a significant effect of impact direction [$\chi^2(1) = 22.63$, $p < .001$], posterior probability [$\chi^2(1) = 8.01$, $p = .005$], and type of evidence [$\chi^2(1) = 18.10$, $p < .001$]. Coherently with the results of Experiment 1, errors were less frequent when the evidence provided confirmed (vs. disconfirmed) the more likely hypothesis and with .6-.4 (vs. .55-.45) posterior. However, across posterior probability levels and regardless of impact direction, errors were more frequent when the type of evidence provided was pattern rather than shape ($M = .45$, SE = .023 and $M = .34$, SE = .019, respectively; post-hoc test: $p < .001$). A significant interaction effect between impact direction and type of evidence was also reported [$\chi^2(1) = 80.67$, $p < .001$]. When the evidence provided was pattern, the frequency of errors was greater when impact and posterior probability were dissociated (*vs* associated), ($M = .64$, SE = .023 and $M = .28$, SE = .032, respectively; post-hoc test: $p < .001$), while such a difference was not observed when the evidence provided was shape, ($p = .916$).

Table 15. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (all trials) in Experiment 3

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,734	,1689	,403	1,065	,000
posterior= .60-.40[1]	-,027	,1463	-,313	,260	,856
posterior= .55-.45[1]	-,088	,1175	-,318	,143	,455
type_of_evidence[2]	-,891	,1720	-1,228	-,554	,000

[1] reference category: .50-.50
[2]reference category: pattern

Table 16. Regression coefficients for the GEE regression on the proportion of errors (all trials) in Experiment 3

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,647	,1039	,443	,850	,000
posterior[1]	-,152	,1873	-,519	,215	,416
type_of_evidence[2]	-1,059	,1563	-1,365	-,752	,000
impact_direction[3]	-1,435	,1774	-1,783	-1,087	,000

[1] reference category: .50-.50
[2]reference category: pattern
[3]reference category: negative

As in Experiments 1 and 2, when the analyses included only consistent trials, results strictly resembled those obtained when all trials were considered. The overall proportion of choices for the confirmed alternative ($M = .61$, SD =.21) was higher than chance level [$t(31) = 2.91$ $p = .007$], and the proportion of errors was greater when the evidence disconfirmed (vs. confirmed) the more likely hypothesis [$M = .48$, SD = .25, and $M = .20$, SD = .24, respectively, $t(31) = 4.17$, $p < .001$]. Type of evidence significantly predicted choices for the confirmed alternatives [$\chi^2(1) = 16.83$, $p < .001$] whereas posterior probability [$\chi^2(2) = .47$, $p = .792$] and the interaction between type of evidence and posterior probability [$\chi^2(2) = 3.39$, $p = .184$] did not. Finally, the GEE regression analysis on errors revealed that impact direction [$\chi^2(1) = 7.92$, $p = .005$], posterior probability [$\chi^2(1) = 5.02$, $p = .025$], and type of evidence [$\chi^2(1) = 8.87$, $p = .003$] were all significant predictors; the only significant interaction effect was between impact direction and type of evidence [$\chi^2(1) = 26.81$, $p < .001$].

Table 17. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (consistent trials) in Experiment 3

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	1,338	,4371	,482	2,195	,002
posterior= .60-.40[1]	-,200	,2633	-,716	,316	,448
posterior= .55-.45[1]	-,168	,1909	-,542	,206	,378
type_of_evidence[2]	-1,618	,4177	-2,436	-,799	,000

[1] reference category: .50-.50
[2]reference category: pattern

Table 18. Regression coefficients for the GEE regression on the proportion of errors (consistent trials) in Experiment 3

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	1,210	,3768	,471	1,948	,001
posterior[1]	-,331	,3251	-,968	,307	,309
type_of_evidence[2]	-1,990	,3481	-2,672	-1,308	,000
impact_direction[3]	-2,645	,7069	-4,030	-1,259	,000

[1] reference category: .50-.50
[2]reference category: pattern
[3]reference category: negative

4.3.5. Discussion

Despite replacing color with another feature, we still found an effect of the feature being provided as evidence. However, type of evidence did not interact significantly with posterior probability in predicting the choice for the confirmed alternative, but it interacted with impact direction in predicting the proportion of errors. As in Experiments 1 and 2, results showed that confirmed alternatives were chosen more frequently than disconfirmed (but equally probable) alternatives, and that evidential impact affected the choice of the normatively correct alternative. Unfortunately, however, the results concerning the choices for confirmed alternative reported again a strong effect of the type of evidence (pattern vs.

shape).

4.4. Experiment 4

4.4.1. Introduction

The replacement of color with pattern aimed at eliminating the asymmetry observed between the two types of evidence employed in Experiments 1 and 2 was unsuccessful, and the results of Experiment 3 showed that also pattern was a more compelling evidence than shape in determining participants' judgments. To the end of contrasting such asymmetry, in Experiment 4, we employed sets of figures that were characterized by the combination of two different features: line curvature (i.e., wavy vs. straight) and line orientation (i.e., horizontal vs. vertical). These two have been suggested as the simplest line features by Treisman and Gormican (1988), and relevant visual search attributes by Wolfe and Horowitz (2017). As in Experiment 2, for each trial, a set of either 23 or 24 figures (i.e., wavy vertical lines, wavy horizontal lines, straight vertical lines, and straight horizontal lines) was presented for 4,000 ms.

4.4.2. Method

Participants

Participants were 32 students of the University of Trento (16% men, mean age= 22.4, *SD= 3).* The study was approved by the ethics committee of the University of Trento, and informed consent was obtained for all participants.

Material and procedure

The procedure was analogous to that described in the previous experiments, with the changes mentioned in the Introduction. Each item of a visual set was composed of lines with different orientation (vertical vs horizontal) and different curvature (wavy vs straight).

4.3.3. Data analysis

The analytical strategy was analogous to that described in Experiment 1.

4.4.4. Results

Table 19. Mean proportions of choice for the confirmed alternative in Experiment 4.

Posterior Probability		All trials	Consistent trials	
	Evidence	Prop.	Prop.	%
0.5-0.5	Orientation			
	Horizontal	.64	.82	47
	Vertical	.65	.73	55
	Curvature			
	Wavy	.53	.60	53
	Straight	.63	.77	58
0.55-0.45	Orientation			
	Horizontal	.66	.73	63
	Vertical	.63	.71	59
	Curvature			
	Wavy	.63	.73	58
	Straight	.59	.59	63
0.6-0.4	Orientation			
	Horizontal	.60	.65	66
	Vertical	.68	.77	73
	Curvature			
	Wavy	.62	.69	63
	Straight	.62	.65	66
Overall		.62	.70	62

"*Prop.*" stands for the mean proportion of choices for the confirmed alternative across participants in all trials (column "*All trials*") or consistent trials only (column "*Consistent trials*"). These proportions have been computed by averaging the proportion of choices for the confirmed option provided by each participant in each subclass of stimuli. "*%*" stands for the percentage of consistent trials out of the total amounts of trials in each subclass of stimuli

Table 20. Mean proportions of errors in Experiment 4

Posterior Probability	Evidence	All trials		Consistent trials			
		Disconfirmed	Confirmed	Disconfirmed		Confirmed	
		Prop.	Prop.	Prop.	%	Prop.	%
0.55-0.45	Orientation						
	Horizontal	.62	.29	.67	64	.17	61
	Vertical	.55	.29	.61	58	.16	61
	Curvature						
	Wavy	.60	.34	.67	48	.26	67
	Straight	.55	.38	.55	70	.29	55
0.6-0.4	Orientation						
	Horizontal	.44	.24	.41	61	.15	70
	Vertical	.53	.16	.45	66	.8	80
	Curvature						
	Wavy	.49	.24	.52	58	.14	67
	Straight	.46	.22	.44	64	.7	69
Overall		.53	.27	.53	61	17	66

"*Prop.*" stands for the mean proportion of errors across participants in trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence, when all trials (column "*All trials*") or only consistent trials (column "*Consistent trials*") are considered. These proportions have been computed by averaging the proportion of errors made by each participant in each subclass of stimuli. "*%*" stands for the percentage of consistent trials out of the total amounts of trials in which the more likely hypothesis was disconfirmed (column "*Disconfirmed*") or confirmed (column "*Confirmed*") by the evidence in each subclass of stimuli.

As in the three previous experiments, the proportion of choices for the confirmed alternative (M = .62, SD = .14) was significantly higher than chance level [$t(31) = 4.98$, $p <$ .001], and the proportion of errors was greater when the most likely hypothesis was disconfirmed (vs. confirmed) by the evidence [M = .53, SD = .18, and M = .27, SD = .14, respectively, $t(31) = 5.23$, $p < .001$]. The GEE regression analysis showed that neither posterior probability nor type of evidence significantly predicted choices for the confirmed alternative [$\chi^2(2) = .78$, $p = .676$, and $\chi^2(1) = 1.79$, $p = .180$, respectively]. The interaction between the two factors was also not significant [$\chi^2(2) = .77$, $p = .679$]. The GEE regression performed on errors revealed that both impact direction and posterior probability were significant predictors [$\chi^2(1) = 26.36$, $p < .001$, and $\chi^2(1) = 28.34$, $p < .001$, respectively], while type of evidence was not [$\chi^2(1) = .83$, $p = .362$]. Errors were less frequent when the evidence provided confirmed (vs. disconfirmed) the most likely hypothesis ($M = .27$, SE = .024 and $M = .53$, SE = .031, respectively; post-hoc test: $p < .001$) and with .6-.4 (vs. .55-.45) posterior ($M = .34$, SE = .017 and $M = .45$, SE = .017, respectively; post-hoc test: $p < .001$). No significant interaction effect was obtained among any of the three considered factors (all $ps > .05$).

Table 21. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (all trials) in Experiment 4

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,595	,1387	,323	,867	,000
posterior= .60-.40[1]	-,009	,1492	-,301	,284	,954
posterior= .55-.45[1]	,017	,1609	-,298	,332	,915
type_of_evidence[2]	-,264	,1592	-,576	,048	,097

[1] reference category: .50-.50
[2]reference category: orientation

Table 22. Regression coefficients for the GEE regression on the proportion of errors (all trials) in Experiment 4

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,347	,1870	-,019	,714	,063
Posterior[1]	-,394	,1991	-,784	-,004	,048
type_of_evidence[2]	-,032	,2657	-,553	,489	,904
impact_direction[3]	-1,247	,3232	-1,881	-,614	,000

[1] reference category: .50-.50
[2]reference category: orientation
[3]reference category: negative

When only consistent trials were included in the analyses, results followed a similar pattern. The proportion of choices for the confirmed alternative ($M = .70$, SD = .23) was significantly higher than chance level [$t(31) = 4.98$, $p < .001$], and the proportion of errors was significantly greater when the evidence disconfirmed (vs. confirmed) the most likely hypothesis [$M = .53$, SD = .24, and $M = .17$, SD = .19, respectively, $t(31) = 5.33$, $p < .001$].

The GEE regression analysis on choices for the confirmed alternative did not reveal any significant effect of posterior probability and type of evidence [$\chi^2(2) = .46$, $p = .794$, and $\chi^2(1) = 2.41$, $p = .121$, respectively], and no interaction between them, [$\chi^2(2) = 2.33$, $p = .312$]. Once again, the GEE regression analysis on errors showed that both impact direction and posterior probability were significant predictors [$\chi^2(1) = 27.09$, $p < .001$, and $\chi^2(1) = 18.72$, $p < .001$, respectively], while type of evidence was not [$\chi^2(1) = .40$, $p = .528$]. No significant interaction effects were obtained among any of the factors considered [all $ps > .05$].

Table 21. Regression coefficients for the GEE regression on the proportion of choices for the confirmed alternative (consistent trials) in Experiment 4

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	1,293	,3375	,631	1,954	,000
posterior= .60-.40[1]	-,416	,3037	-1,011	,179	,171
posterior= .55-.45[1]	-,228	,3525	-,919	,463	,518
type_of_evidence[2]	-,683	,3343	-1,338	-,028	,041

[1] reference category: .50-.50
[2]reference category: orientation

Table 22. Regression coefficients for the GEE regression on the proportion of errors (consistent trials) in Experiment 4

Parameter	B	Std. Error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
(Intercept)	,580	,3172	-,042	1,201	,068
Posterior[1]	-,654	,3304	-1,301	-,006	,048
type_of_evidence[2]	-,041	,4662	-,955	,873	,930
impact_direction[3]	-2,285	,5861	-3,433	-1,136	,000

[1] reference category: .50-.50
[2]reference category: orientation
[3]reference category: negative

4.4.5. Discussion

In line with our hypotheses, participants chose the confirmed alternative over the disconfirmed one across posterior probability levels and types of evidence. The proportion of errors was significantly higher when impact and posterior probability were dissociated, across posterior probability levels and types of evidence. The combination of line curvature and orientation proved to be the more balanced among those employed in the present research. Only in this last experiment, indeed, the type of evidence did not affect the choice for the confirmed alternative, nor the amount of errors and no interaction effects between the independent variables were found. When the two perceptual features at issue are balanced, relative posterior probability of the two hypotheses does not interact with the type of evidence being provided when having to choose the confirmed alternative. Similarly, impact direction did affect the proportion of errors in participants' choices regardless of the relative posterior probability of the two hypotheses and of the type of evidence being tested.

Chapter 5: General discussion, further directions and conclusions

The goal of this work was to investigate whether confirmation relations affect probability judgments even when abstract, perceptual material is involved. To test this hypothesis, we ran four behavioral experiments in which we used the same paradigm and manipulated perceptual features. Overall, the results we found supported our experimental claims, showing that probability judgments are affected by evidential impact even when perceptual stimuli are at issue. However, if the perceptual features involved were characterized by different salience levels, this asymmetry affected probability judgments, as shown in Experiments 1 to 3. When the evidence provided was the more salient feature, participants' choices aligned with our hypotheses: confirmed alternative was chosen over the disconfirmed one, and the proportion of errors was significantly higher when probability and impact diverged. On the other hand, when the less salient feature was given as evidence, participants were virtually insensitive to variations in impact and probability: their choice for the confirmed alternative across posterior probability levels would generally not deviate from chance level. Insensitivity to impact and probability differences also affected the percentages of errors: with the less salient feature as evidence, in fact, the percentage of errors did not differ in the two levels of the confirmation/probability relation. In Experiment 4, the two perceptual features composing each figure seemed to be more balanced; in this last case, choice patterns were not affected by the feature provided as evidence, nor by the relative posterior probability of the two hypotheses, as in Experiments 1 to 3. In the last experiment, in fact, participants chose the confirmed alternative over the disconfirmed one regardless of the posterior probability of the two hypotheses and of the type of evidence provided. Consequently, percentages of errors were significantly higher when the more probable hypothesis was disconfirmed rather than confirmed. This difference was consistent across posterior probability levels as well as types of evidence.

In addition to supporting our hypotheses, these results suggest the existence of an influence from confirmation relations between perceptual features on probability judgments, which is relevant because it shows that confirmation relations are capable of affecting probability judgments even in absence of any semantic element. In fact, past works on confirmation relations involved material with semantic background: it is possible that it is easier to think about confirmation relations and process them when they involve concrete features that can be easily represented, whereas confirmation relations between abstract, geometrical features appear much more arbitrary and hard to process. Another downside of abstract features, as suggested by our results, is that they strongly affected participants' responses in three out of four experiments. In fact, when the more salient feature (color vs. shape, pattern vs shape) was provided as evidence, participants were able to correctly choose the confirmed alternative between two options; on the other hand, when the less salient feature appeared as evidence, people appeared to be insensitive to the confirmation relations pointing to one or the other alternative. A possible explanation for this finding is that people focus on the more salient of the two features composing each item and organize the visual set in smaller subsets based on it. Therefore, when this feature is given as evidence, people have a clear idea of the conditional probabilities it entails. On the other hand, when the less salient feature is given as evidence, people's choices for the confirmed alternative will not differ from chance level and they also will not be affected by the evidence confirming or disconfirming the normatively correct alternative when asked to choose the more probable one.

Moreover, despite not reporting any significant effect of participants' gender on choice pattern, the samples of all four experiments are unbalanced with respect to gender: this could represent one possible limitation of the study.

Our experimental results can have alternative explanations which fall outside the confirmation-driven framework. As to Experiments 1 and 2, the asymmetry in salience between shape and color of each item could have affected choice patterns regardless of the statistical structure of each set; that is, the more salient feature could have captured participant's attention and driven their choices, regardless of the other one. To try solving this issue, in Experiment 3 we removed the feature 'color' and replaced it with 'pattern'. Yet again, even with this new visual set we obtained, shape was overlooked in favor of pattern. The explanation we propose for such asymmetry is analogous to the one involving color: the more salient feature overcame the less salient one in driving participants' choices because it would allow people to form subsets of the visual set, as mentioned above. Another alternative account could rely mostly, or exclusively, on the visual features of the item(s) which appeared in the center of the visual set. For example: if participants saw a white circle in the center of the screen, they could focus their attention on this item and perceive a stronger relation between the two features than the actual one and this judgment would then affect the choice in the experimental task. We randomized the position of all the items of a set on the screen to minimize the probability of incurring in such a position-driven salience effect, but this is still a possible explanation for participants' choices. However, we believed that this effect would disappear when averaging throughout the entire sample.

Our experimental results have theoretical as well as applied implications. On a theoretical level, they extend results coming from works involving verbal and linguistic material (Paperno et al., 2014; Tentori, Chater & Crupi, 2016) to perceptual stimuli with no semantic background and provide further evidence for the effect of impact relations on probability judgments. Additionally, they show that high-level relations, which are completely unknown to the subject, affect the way people perceive relations within a visual set of perceptual, two-dimensional items. This last result might have interesting and noteworthy implications for psychophysical studies on visual cognition and contingency learning because it shows how non-perceptual relations between items might affect those perceptual processes on which visual cognition and search are based on. Finally, the asymmetry we found in choice patterns in experiments 1 to 3 reflects evidence coming from visual perception and search studies (Wolfe & Horowitz, 2004, 2017), revealing how color and shape are not equally salient. To understand this asymmetry, a more in-depth discussion of color and shape effects in visual perception is due.

5.1 Color, shape and other features in visual search

Visual attention can be driven to different extent by different perceptual features: Treisman & Gormican (1988) reviewed works involving different perceptual features, like color, line length, curvature, orientation, and arrangements, and used search latencies to infer which features are coded automatically in early vision. Kaptein, Theeuwes, & van der Heijden (1995) ran four experiments to investigate to what extent subjects are capable of selectively limiting search to a subset of elements based on color. The results showed that when subjects were searching for a target defined as a conjunction of color and orientation, response latencies on target-present trials depended only on the number of elements in the non-target color, indicating that only the subset of elements in the target color was searched.

Color based subsets, then, seem to have a crucial role in guiding visual attention. The importance of color in the deployment of visual attention is further discussed in a more recent work (Adamo et al., 2010) aimed at comparing the relative efficacy of shape and color in modulating attentional capture. The results provided evidence of attentional capture by both color and shape, but the magnitude of the effect was significantly larger for color than for shape. These two features, then, might be associated with different levels within the visual processing hierarchy and sets defined by color may be applied more effectively than those defined by shape: on a concrete level, this meant that in the experiment participants were more likely to orient attention to an irrelevant shape cue than to an irrelevant color cue. An additional evidence of the relevance of color in visual search can be found in several works by Wolfe (Wolfe & Horowitz, 2004, 2017) aimed at investigating what attributes guide the deployment of attention in visual search. These studies suggested that some features are more powerful than others in leading visual attention or, in other terms, that there are some attributes that undoubtedly guide the deployment of attention, like color or orientation, and some others for which there is no such strong support, like color change or semantic category.

Color and shape can also prime visual perception to different extent: Breitmeyer, Ogmen and Chen (2004) performed a psychophysical and neurophysiological investigation of the types and levels of unconscious processing in color and form perception: the joint results from psychophysical and neurophysiological analyses showed that color priming happens earlier than shape priming. A series of experiments focused on both color and form indicated that choice reaction times for both color and shape are affected by priming in a metacontrast masking paradigm. According to the authors, these results indicate that color and form priming occur at different levels in the visual processing stream; color priming depends on a stimulus-dependent response coming from early cortical levels, whereas form priming occurs at later levels. An additional experimental work from the same author (Breitmeyer, Ro, & Singhal, 2004) provided further proof of an early unconscious color priming effect occurring at early levels of processing. This second research involved two behavioral, computer-based experiments aimed at exploring unconscious priming effect of differently colored stimuli. This asymmetry between color and shape priming could be extremely relevant for experiments where shape and color are the two critical dimensions.

5.2. Further steps

Several further steps could help to better understand the conditions under which confirmation and probability interact. One could involve presentation modality: given the large number of items presented in each visual set, subitizing was not an option; however, one could argue that with a simultaneous presentation, people might be susceptible to perceptual grouping and item locations (we obviously randomized object locations in our experiments). Sequential presentation could represent a viable option to control for these issues: all the items of a same set could be shown sequentially instead of making up a visual set. This different modality is not without critical points either: showing items in a sequence might give rise to some primacy or recency effects. Ideally, the two modalities could counterbalance each other's flaws.

The studies we presented demonstrated that, even when performing probabilistic inferences, people are affected by impact relations in their judgments. It would be interesting to explore how people perceive and conceptualize confirmation relations when explicitly asked about them, that is, what they would base their judgments on when the focus is on confirmation. This would allow to test whether and how people process and assess confirmation relations when explicitly asked to do so, instead of inferring their strategy from their behavior in a probability judgment task. The evidence collected so far in the psychological literature showed that probability estimates depend on confirmation relations; evidence that impact relations do not depend on probabilities would further support the confirmation-theoretic account. To explore choice patterns in impact judgment tasks, it is crucial to conceptualize confirmation relations so that people can understand and subsequently assess them. When semantic content is available, the relation between hypothesis and evidence is transparent and easy to conceptualize, but it gets much more blurry and arbitrary in presence of perceptual stimuli with no intrinsic meaning and link. In fact, very good performance at estimating impact relations does not necessarily entail that people would find it as easy to conceptualize them. This is particularly difficult when abstract stimuli are at issue: while it is relatively easy to think of how two real-life concepts are associated to each other and to grasp the difference between confirmation and probability relations between them (as in Tentori et al., 2016), this difference is much less intuitive when abstract, geometric shapes are involved and their relations are purely arbitrary and not based on any background knowledge. Moreover, confirmation relations are relative, not absolute judgments: they capture the relation between two variables and the (positive or negative) change in someone's belief. Therefore, to investigate them in a direct way one would need a reference point to depart from after assessing the impact of the evidence; such an operationalization, though, would defy the whole idea of confirmation being a more primitive notion than probability. From a purely theoretical point of view, other concepts like contingency or association measures could be considered as proxies for this notion, as already discussed in Chapter 2. An experimental question focused on confirmation instead of probability judgments would allow to understand whether the former is psychologically prior to the latter, as proposed by Tentori, Crupi, Bonini & Osherson (2007). If this were the case, confirmation relations, and not probability ones, might be the basis of probabilistic inference.

In Chapter 3, I described the so-called Bayesian brain hypothesis and its implications for high- and low-level cognitive processes. Two central claims of this chapter were that the brain represents sensory information probabilistically and that it is a Bayesian sampler working with a local sense of probability. Our results suggest that confirmation relations might play a crucial role in this suboptimal but effective reasoning process.

As a side note, our results could inform the debate focused on cognitive penetrability. Cognitive penetration "describes the influence of higher level cognitive factors on perceptual experience" (Vetter & Newen, 2014) and implies that high-level (in our case, confirmation) relations affect the perception of some other high-level (probability) relations between low-level (perceptual) features. We could think of our experimental results as a very particular instance of cognitive penetrability where the high-level (confirmation) relations could 'penetrate' the elaboration of the perceptual, low-level material.

More and less recent experimental works showed distinct brain loci for deductive versus probabilistic reasoning (Osherson et al., 1998) and for priors versus likelihoods encoding (Vilares et al., 2012; Wiener, Michaelis, & Thompson, 2016). More precisely, neuroimaging studies showed that, in perceptual and cognitive decision-making tasks, priors and likelihoods are encoded separately, and so are priors and likelihood related- uncertainty.

For example, Ting, Yu, Maloney, & Wu (2015) were able to identify priors and likelihoods as distinct sources of information in value-based decision making, and localize prior-likelihood integration in the brain. Similarly, Ma & Jazayeri (2014) found differential representations of prior and likelihood uncertainty in the human brain, meaning that humans can take both types of uncertainty into account in their computations. Further investigations could explore whether probability estimates and confirmation judgments are also encoded in different brain areas, that is, whether a conceptual dissociation between the two notions is associated to a spatial and/or functional separation. More specifically: neuroimaging techniques could explore whether activation patterns in probabilistic inferences resemble more those associated to priors and likelihoods encoding or whether they are characterized by a whole different pattern. If the first instance were true, it would provide evidence for probability-driven inferences; otherwise, it would further support the confirmation-driven account.

In this work, we demonstrated that confirmation relations affect probability judgments even when abstract stimuli with no semantic content are at issue. If people are sensitive to arbitrary confirmation relations, it is possible to hypothesize that they are even more sensitive to impact relations between semantically charged concepts, as, indeed, demonstrated by empirical works we discussed in Chapter 2. In the same chapter, we discussed one possible consequence of the sensitivity to confirmation relations, that is conjunction fallacy. This fallacy was found and demonstrated in controlled, experimental settings but can nonetheless affect probabilistic judgments in social settings leading, for example, to stereotypical judgment. Similar to reasoning fallacies and heuristics, stereotypes provide a way to make sense of the complexity in the surrounding environment. Under most circumstances, they help us behave (i.e. think, take decisions, interact) in an adaptive and sound manner, but in some situations, they can lead to suboptimal judgment and behavior. Social psychologists proposed several explanations for stereotypes, mostly focused on motivational factors: people engage in stereotyped judgments to make sense of the social environment, protect their self (*self-serving bias*) and their social identity (*social identity threat*). Some other explanations relate to cognitive factors: Kutzner & Fiedler (2017) proposed an explanatory account for stereotypes based on pseudocontingencies between events or features. These pseudocontingencies lead people to assume correlations between attributes based solely on differences between base rates. The authors defined stereotypes as "subjectively expected statistical contingencies between attributes and social groups" which, then, could be quantified by ΔP. Environmental factors, such as rarity, proximity, group categorization make pseudocontingencies and illusory correlations easier to draw. In this thesis, I showed that probability judgment is affected by confirmation relations and proposed a parallel between confirmation and contingency relations. In stereotypical judgments, people learn to associate personality traits and social categories on the basis of their experience; this association is not based on probabilistic assumptions, but on the observation that belonging to a given social group seems to 'confirm' the presence of a certain personal trait. Despite allowing quick and effortless social judgments, stereotypical relations do not necessarily represent the environment in a truthful and accurate way. Similarly, confirmation relations often represent an efficient and easily retrieved tool for probability estimates despite representing a deviation from the normative benchmark. If, as proposed by Kutzner & Fiedler (2017), stereotypes are based on pseudocontingencies between events or features, then it is possible to hypothesize that confirmation-driven judgments could be at the core of stereotypical associations.

References

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS computational biology*, *10*(6), e1003661. https://doi.org/10.1371/journal.pcbi.1003661

Adamo, M., Wozny, S., Pratt, J., & Ferber, S. (2010). Parallel, independent attentional control settings for colors and shapes. *Attention, Perception, & Psychophysics*, *72*(7), 1730-1735. https://doi.org/10.3758/APP.72.7.1730

Allan, L. G. (1980). A note on measurement of contingency between two binary variables in judgment tasks. *Bulletin of the Psychonomic Society*, *15*(3), 147–149. https://doi.org/10.3758/BF03334492

Allan, L. G., Hannah, S. D., Crump, M. J., & Siegel, S. (2008). The psychophysics of contingency assessment. *Journal of Experimental Psychology: General*, *137*(2), 226-243. http://dx.doi.org/10.1037/0096-3445.137.2.226

Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation*, *14*(4), 381–405. https://doi.org/10.1016/0023-9690(83)90024-3

Allan, L. G., & Siegel, S. (2002). A signal detection theory analysis of the placebo effect. *Evaluation & the Health Professions*, *25*(4), 410-420. https://doi.org/10.1177/0163278702238054

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological science*, *12*(2), 157-162. https://doi.org/10.1111%2F1467-9280.00327

Baker, A. G., Berbrier, M. W., & Vallee-Tourangeau, F. (1989). Judgements of a 2× 2 contingency table: Sequential processing and the learning curve. *The Quarterly Journal of Experimental Psychology Section B*, *41*(1b), 65-97. https://doi.org/10.1080%2F14640748908401184

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. Acta Psychologica, 44(3), 211–233. https://doi.org/10.1016/0001-6918(80)90046-3

Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., ... & Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142-1152. https://doi.org/10.1016/j.neuron.2008.09.021

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30-39. https://doi.org/10.1016/j.neuron.2012.03.016

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of vision*, *9*(5), 23-23. https://doi.org/10.1167/9.5.23

Bowers, J. S., & Davis, C. J. (2012a). Bayesian just-so stories in psychology and neuroscience. *Psychological bulletin*, *138*(3), 389-414. http://dx.doi.org/10.1037/a0026450

Bowers, J. S., & Davis, C. J. (2012b). Is that what Bayesians believe? reply to Griffiths, Chater, Norris, and Pouget (2012). https://doi.org/10.1037/a0027750

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433-436. https://doi.org/10.1163/156856897X00357

Breitmeyer, B. G., Ogmen, H., & Chen, J. (2004). Unconscious priming by color and form: Different processes and levels. *Consciousness and cognition*, *13*(1), 138-157. https://doi.org/10.1016/j.concog.2003.07.004

Breitmeyer, B. G., Ro, T., & Singhal, N. S. (2004). Unconscious color priming occurs at stimulus-not percept-dependent levels of processing. *Psychological Science*, *15*(3), 198-202. https://doi.org/10.1111%2Fj.0956-7976.2004.01503009.x

Carnap, R. (1950). *Logical foundations of probability.*

Chan, S. C., Niv, Y., & Norman, K. A. (2016). A probability distribution over latent causes, in the orbitofrontal cortex. *Journal of Neuroscience*, *36*(30), 7817-7828. https://doi.org/10.1523/JNEUROSCI.0659-16.2016

Chapman, G. B., & Robbins, S. J. (1990). Cue interaction in human contingency judgment. *Memory & Cognition*, *18*(5), 537-545. https://doi.org/10.3758/BF03198486

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. https://doi.org/10.1016/j.tics.2006.05.007

Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 811-823. https://doi.org/10.1002/wcs.79

Cherubini, P. (2005). *Psicologia del pensiero*. Raffaello Cortina.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta psychologica*, *138*(2), 289-301. https://doi.org/10.1016/j.actpsy.2011.08.002

Crump, M. J. C., Hannah, S. D., Allan, L. G., & Hord, L. K. (2007). Contingency judgements on the fly. *Quarterly Journal of Experimental Psychology (2006)*, *60*(6), 753–761. https://doi.org/10.1080/17470210701257685

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, *14*(2), 182-199. https://doi.org/10.1080/13546780701643406

Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, *74*(2), 229-252. https://doi.org/10.1086/520779

De Ridder, D., Vanneste, S., & Freeman, W. (2014). The Bayesian brain: phantom percepts resolve sensory uncertainty. *Neuroscience & Biobehavioral Reviews*, *44*, 4-15. https://doi.org/10.1016/j.neubiorev.2012.04.001

Domurat, A., Kowalczuk, O., Idzikowska, K., Borzymowska, Z., & Nowak-Przygodzka, M. (2015). Bayesian probability estimates are not necessary to make choices satisfying

Bayes' rule in elementary situations. *Frontiers in psychology*, *6*, 1194. https://doi.org/10.3389/fpsyg.2015.01194

Eells, E., & Fitelson, B. (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies*, *107*(2), 129-142. https://doi.org/10.1023/A:1014712013453

Elqayam, S., & Evans, J. S. B. (2011). Subtracting "ought" from "is": Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, *34*(5), 233-248. https://doi.org/10.1017/S0140525X1100001X

Feldman, J. (2014). Bayesian models of perceptual organization. *Handbook of perceptual organization*. https://doi.org/10.1093/oxfordhb/9780199686858.013.007

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, 294–299. 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*(3), 415-22. http://dx.doi.org/10.1037/a0026884

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. https://doi.org/10.1017/CBO9780511816772.006

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological science*, *17*(9), 767-773. https://doi.org/10.1111/j.1467-9280.2006.01780.x

Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *Oxford* *series in visual cognition. From perception to consciousness: Searching with Anne Treisman* (pp. 339-349). New York, NY, US: Oxford University Press. http://dx.doi.org/10.1093/acprof:osobl/9780199734337.003.0030

Hahn, U. (2014). The Bayesian boom: good thing or bad? *Frontiers in psychology*, *5*, 765. https://doi.org/10.3389/fpsyg.2014.00765

Hannah, S. D., Crump, M. J., Allan, L. G., & Siegel, S. (2009). Cue-interaction effects in contingency judgments using the streamed-trial procedure. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *63*(2), 103-112. https://doi.org/10.1037/a0013521

Hartmann, S., & Sprenger, J. (2010). Bayesian epistemology. *Routledge companion to epistemology*, 609-620.

Hawthorne, J. (2011). Confirmation theory. In Prasanta S. Bandyopadhyay & Malcolm Forster (eds.), *Philosophy of Statistics, Handbook of the Philosophy of Science, Volume 7*. Elsevier.

Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley interdisciplinary reviews: Cognitive science*, *1*(2), 278-292. https://doi.org/10.1002/wcs.44

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. *Rational models of cognition*, 248-274.

Helmholtz, H. (1878). The facts of perception. *Selected Writings of Hermann Helmholtz*, 1-15.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision research*, *49*(10), 1295-1306. https://doi.org/10.1016/j.visres.2008.09.007

Jarvstad, A., Hahn, U., Warren, P. A., & Rushton, S. K. (2014). Are perceptuo-motor decisions really more optimal than cognitive decisions? Cognition, 130(3), 397–416. https://doi.org/10.1016/j.cognition.2013.09.009

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169-188. https://doi.org/10.1017/S0140525X10003134

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological review*, *116*(4), 856-74. https://doi.org/10.1037/a0016979

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237. http://dx.doi.org/10.1037/h0034747

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition*, *11*(2), 123-141. http://psycnet.apa.org/doi/10.1016/0010-0277(82)90022-1

Kaptein, N. A., Theeuwes, J., & Van der Heijden, A. H. C. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(5), 1053-1069. http://dx.doi.org/10.1037/0096-1523.21.5.1053

Kemeny, J. G., & Oppenheim, P. (1952). Degree of factual support. *Philosophy of Science*, *19*(4), 307-324. https://doi.org/10.1086/287214

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*(1), 20–58. https://doi.org/10.1037/a0014282

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annu. Rev. Psychol.*, *55*, 271-304. https://doi.org/10.1146/annurev.psych.55.090902.142005

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712-719. https://doi.org/10.1016/j.tins.2004.10.007

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in cognitive sciences*, *10*(7), 319-326. https://doi.org/10.1016/j.tics.2006.05.003

Kutzner, F., & Fiedler, K. (2017). Stereotypes as pseudocontingencies. *European Review of Social Psychology*, *28*(1), 1-49. https://doi.org/10.1080/10463283.2016.1260238

Laquitaine, S., & Gardner, J. L. (2018). A Switching Observer for Human Perceptual Estimation. *Neuron*, *97*(2), 462-474. https://doi.org/10.1016/j.neuron.2017.12.011

Leshinskaya, A., & Thompson-Schill, S. L. (2018). Inferences about Uniqueness in Statistical Learning. *Behavioral and Brain Sciences*, *32*(02), 183.

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual review of neuroscience*, *37*, 205-220. https://doi.org/10.1146/annurev-neuro-071013-014017

Mamassian, P., Landy, M., & Maloney, L. (2002). Bayesian modeling of visual perception. In R. Rao, M. Lewicki, & B. Olshausen (Eds.), *Probabilistic models of the brain: Perception and neural function* (pp. 13-36). Cambridge, MA: MIT Press.

Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General*, *127*(3), 269-285. http://dx.doi.org/10.1037/0096-3445.127.3.269

Manktelow, K. I. (2012). *Thinking and reasoning*. Hove, UK: Psychology Press

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*(1), 33–61. https://doi.org/10.1016/j.cogpsych.2006.04.004

Meng, M., & Tong, F. (2004). Can attention selectively bias bistable perception? Differences between binocular rivalry and ambiguous figures. *Journal of vision*, *4*(7), 2-2. https://dx.doi.org/10.1167%2F4.7.2

Moreno-Bote, R., Knill, D. C., & Pouget, A. (2011). Bayesian sampling in visual perception. *Proceedings of the National Academy of Sciences*, *108*(30), 12491-12496. https://doi.org/10.1073/pnas.1101430108

Nozick, R. (1981). *Philosophical explanations*. Harvard University Press.

Osherson, D., Perani, D., Cappa, S., Schnur, T., Grassi, F., & Fazio, F. (1998). Distinct brain loci in deductive versus probabilistic reasoning. Neuropsychologia, 36(4), 369–376. https://doi.org/10.1016/S0028-3932(97)00099-7

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

Pelli, D. G., & Vision, S. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437-442. https://doi.org/10.1163/156856897X00366

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature neuroscience*, *16*(9), 1170. https://doi.org/10.1038/nn.3495

Prime, H., & Shultz, T. (2011). Explicit Bayesian reasoning with frequencies, probabilities, and surprisals. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).

Rescorla, M. (2015). Bayesian perceptual psychology. *The Oxford handbook of the philosophy of perception*, 694-716. https://doi.org/10.1093/oxfordhb/9780199600472.013.010

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in cognitive sciences*, *20*(12), 883-893. https://doi.org/10.1016/j.tics.2016.10.003

Sanjana, N. E., & Tenenbaum, J. B. (2003). Bayesian models of inductive generalization. In *Advances in neural information processing systems* (pp. 59-66).

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in human neuroscience*, *7*, 668. https://doi.org/10.3389/fnhum.2013.00668

Shanks, D. R. (1985). Continuous monitoring of human contingency judgment across trials. Memory & Cognition, 13(2), 158–167. https://doi.org/10.3758/BF03197008

Sides, A., Osherson, D., Bonini, N. N. N., & Viale, R. (2002). On the reality of the conjunction fallacy. *Memory and Cognition*, *30*(2), 191–198. https://doi.org/10.3758/BF03195280

Sirota, M., Vallée-Tourangeau, G., Vallée-Tourangeau, F., & Juanchich, M. (2015). On Bayesian problem-solving: helping Bayesians solve simple Bayesian word problems. *Frontiers in psychology*, *6*, 1141. https://doi.org/10.3389/fpsyg.2015.01141

Sloman, S. A., & Lagnado, D. (2005). The problem of induction. *The Cambridge handbook of thinking and reasoning*, 95-116.

Spratling, M. W. (2016). A neural implementation of Bayesian inference based on predictive coding. *Connection Science*, *28*(4), 346-383. https://doi.org/10.1080/09540091.2016.1243655

Tapia, E., Breitmeyer, B. G., & Shooner, C. R. (2010). Role of Task-Directed Attention in Nonconscious and Conscious Response Priming by Form and Color, 36(1), 74–87. https://doi.org/10.1037/a0017166

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in cognitive sciences*, *10*(7), 309-318. https://doi.org/10.1016/j.tics.2006.05.009

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*(3), 758-778. https://doi.org/10.1111/cogs.12259

Tentori, K., Crupi, V., Bonini, N., & Osherson, D. (2007). Comparison of confirmation measures. *Cognition*, *103*(1), 107-119. https://doi.org/10.1016/j.cognition.2005.09.006

Tentori, K., Crupi, V., & Russo, S. (2013). On the determinants of the conjunction fallacy: Probability versus inductive confirmation. *Journal of Experimental Psychology: General*, *142*(1), 235-255. https://doi.org/10.1037/a0028770

Ting, C. C., Yu, C. C., Maloney, L. T., & Wu, S. W. (2015). Neural mechanisms for integrating prior knowledge and likelihood in value-based probabilistic inference. *Journal of Neuroscience*, *35*(4), 1792-1805. https://doi.org/10.1523/JNEUROSCI.3161-14.2015

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: evidence from search asymmetries. *Psychological review*, *95*(1), 15-48. http://doi.org/10.1037/0033295X.95.1.15

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A., & Kahneman, D. (1981). *Judgments of and by representativeness* (No. TR-3). STANFORD UNIV CA DEPT OF PSYCHOLOGY.

Utochkin, I. S., Khvostov, V. A., & Stakina, Y. M. (2018). Continuous to discrete: Ensemble-based segmentation in the perception of multiple feature conjunctions. *Cognition*, *179*, 178-191. https://doi.org/10.1016/j.cognition.2018.06.016

Vadillo, M. A., Miller, R. R., & Matute, H. (2005). Causal and predictive-value judgments, but not predictions, are based on cue-outcome contingency. *Learning & Behavior*, *33*(2), 172-183. https://doi.org/10.3758/BF03196061

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and cognition*, *27*, 62-75. https://doi.org/10.1016/j.concog.2014.04.007

Vilares, I., Howard, J. D., Fernandes, H. L., Gottfried, J. A., & Körding, K. P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Current Biology*, *22*(18), 1641-1648. https://doi.org/10.1016/j.cub.2012.07.010

Vossel, S., Mathys, C., Stephan, K. E., & Friston, K. J. (2015). Cortical coupling reflects Bayesian belief updating in the deployment of spatial attention. *Journal of Neuroscience*, *35*(33), 11532-11542. https://doi.org/10.1523/JNEUROSCI.1382-15.2015

Wei, X. X., & Stocker, A. A. (2012). Bayesian inference with efficient neural population codes. In *International Conference on Artificial Neural Networks* (pp. 523-530). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33269-2_66

Wiener, M., Michaelis, K., & Thompson, J. C. (2016). Functional Correlates of Likelihood and Prior Representations in a Virtual Distance Task, 3187, 3172–3187. https://doi.org/10.1002/hbm.23232

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*(6), 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*(3), 0058. https://doi.org/10.1038/s41562-017-0058

Yang, Y., Tokita, M., & Ishiguchi, A. (2018). Is there a common summary statistical process for representing the mean and variance? A study using illustrations of familiar items. *i-Perception*, *9*(1). https://doi.org/10.1177%2F2041669517747297

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis?. *Trends in cognitive sciences*, *10*(7), 301-308. https://doi.org/10.1016/j.tics.2006.05.002

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological science*, *24*(5), 667-677. https://doi.org/10.1177%2F0956797612460407

Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. B. (2011). Mutual interference between statistical summary perception and statistical learning. *Psychological Science*, *22*(9), 1212-1219. https://doi.org/10.1177/0956797611419304

Zhao, J., Shah, A., & Osherson, D. (2009). On the provenance of judgments of conditional probability. *Cognition*, *113*(1), 26-36. https://doi.org/10.1016/j.cognition.2009.07.006

Zhao, J., & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, *146*, 217-222. https://doi.org/10.1016/j.cognition.2015.09.018

Appendix: experimental sets used in Experiment 1.

Figure 3. Sets 1 to 10. Evidence: shape (triangle); hypothesis: color.

1

	white	black		priors
triangle	12	12	24	0.50
circle	20	4	24	0.50
	32	16	**48**	
priors	0.667	0.333		

Pr(W|triangle)= 0.50 L(W, triangle)= -0.33

Pr(B|triangle)= 0.50 L(B, triangle) = 0.33

2

	white	black		priors
triangle	12	12	24	0.50
circle	4	20	24	0.50
	16	32	**48**	
priors	0.333	0.667		

Pr(W|triangle) 0.50 L(W, triangle)= 0.33

Pr(B|triangle) 0.50 L(B, triangle) = -0.33

3

	white	black		priors
triangle	12	10	22	0.48
circle	22	2	24	0.52
	34	12	**46**	
priors	0.739	0.261		

Pr(W|triangle)= 0.55 L(W, triangle)= -0.40

Pr(B|triangle)= 0.45 L(B, triangle) = 0.40

4

	white	black		priors
triangle	12	10	22	0.48
circle	4	20	24	0.52
	16	30	**46**	
priors	0.348	0.652		

Pr(W|triangle) 0.55 L(W, triangle)= 0.38

Pr(B|triangle) 0.45 L(B, triangle) = -0.38

5

	black	white		priors
triangle	12	10	22	0.48
circle	22	2	24	0.52
	34	12	**46**	
priors	0.739	0.261		

Pr(B|triangle)= 0.55 L(B, triangle) = -0.40

Pr(W|triangle)= 0.45 L(W, triangle)= 0.40

6

	black	white		priors
triangle	12	10	22	0.48
circle	4	20	24	0.52
	16	30	**46**	
priors	0.348	0.652		

Pr(B|triangle) 0.55 L(B, triangle) = 0.38

Pr(W|triangle) 0.45 L(W, triangle)= -0.38

7

	white	black		priors
triangle	12	8	20	0.43
circle	24	2	26	0.57
	36	10	**46**	
priors	0.783	0.217		

Pr(W|triangle)= 0.60 L(W, triangle)= -0.41

Pr(B|triangle)= 0.40 L(B, triangle) = 0.41

8

	white	black		priors
triangle	12	8	20	0.43
circle	6	20	26	0.57
	18	28	**46**	
priors	0.391	0.609		

Pr(W|triangle) 0.60 L(W, triangle)= 0.40

Pr(B|triangle) 0.40 L(B, triangle) = -0.40

9

	black	white		priors
triangle	12	8	20	0.43
circle	24	2	26	0.57
	36	10	**46**	
priors	0.783	0.217		

Pr(B|triangle)= 0.60 L(B, triangle) = -0.41

Pr(W|triangle)= 0.40 L(W, triangle)= 0.41

10

	black	white		priors
triangle	12	8	20	0.43
circle	6	20	26	0.57
	18	28	**46**	
priors	0.391	0.609		

Pr(B|triangle) 0.60 L(B, triangle) = 0.40

Pr(W|triangle) 0.40 L(W, triangle)= -0.40

Figure 4. Sets 11 to 20. Evidence: shape (circle); hypothesis: color

11

	white	black		priors
circle	12	12	24	0.50
triangle	20	4	24	0.50
	32	16	**48**	
priors	0.667	0.333		

Pr(W|circle)= 0.50 L(W, circle)= -0.33

Pr(B|circle)= 0.50 L(B, circle) = 0.33

12

	white	black		priors
circle	12	12	24	0.50
triangle	4	20	24	0.50
	16	32	**48**	
priors	0.333	0.667		

Pr(W|circle)= 0.50 L(W, circle)= 0.33

Pr(B|circle)= 0.50 L(B, circle) = -0.33

13

	white	black		priors
circle	12	10	22	0.48
triangle	22	2	24	0.52
	34	12	**46**	
priors	0.739	0.261		

Pr(W|circle)= 0.55 L(W, circle)= -0.40

Pr(B|circle)= 0.45 L(B, circle) = 0.40

14

	white	black		priors
circle	12	10	22	0.48
triangle	4	20	24	0.52
	16	30	**46**	
priors	0.348	0.652		

Pr(W|circle)= 0.55 L(W, circle)= 0.38

Pr(B|circle)= 0.45 L(B, circle) = -0.38

15

	black	white		priors
circle	12	10	22	0.48
triangle	22	2	24	0.52
	34	12	**46**	
priors	0.739	0.261		

Pr(B|circle)= 0.55 L(B, circle) = -0.40

Pr(W|circle)= 0.45 L(W, circle)= 0.40

16

	black	white		priors
circle	12	10	22	0.48
triangle	4	20	24	0.52
	16	30	**46**	
priors	0.348	0.652		

Pr(B|circle)= 0.55 L(B, circle) = 0.38

Pr(W|circle)= 0.45 L(W, circle)= -0.38

17

	white	black		priors
circle	12	8	20	0.43
triangle	24	2	26	0.57
	36	10	**46**	
priors	0.783	0.217		

Pr(W|circle)= 0.60 L(W, circle)= -0.41

Pr(B|circle)= 0.40 L(B, circle) = 0.41

18

	white	black		priors
circle	12	8	20	0.43
triangle	6	20	26	0.57
	18	28	**46**	
priors	0.391	0.609		

Pr(W|circle)= 0.60 L(W, circle)= 0.40

Pr(B|circle)= 0.40 L(B, circle) = -0.40

19

	black	white		priors
circle	12	8	20	0.43
triangle	24	2	26	0.57
	36	10	**46**	
priors	0.783	0.217		

Pr(B|circle)= 0.60 L(B, circle) = -0.41

Pr(W|circle)= 0.40 L(W, circle)= 0.41

20

	black	white		priors
circle	12	8	20	0.43
triangle	6	20	26	0.57
	18	28	**46**	
priors	0.391	0.609		

Pr(B|circle)= 0.60 L(B, circle) = 0.40

Pr(W|circle)= 0.40 L(W, circle)= -0.40

Figure 5. Sets 21 to 30. Evidence: color (black); hypothesis: shape

21	triangle	circle		priors	$Pr(T \mid black)=$	0.50	L(T, black)=	-0.33
black	12	12	24	0.50				
white	20	4	24	0.50	$Pr(C \mid black)=$	0.50	L(C, black) =	0.33
	32	16	**48**					
priors	0.667	0.333						

22	triangle	circle		priors	$Pr(T \mid black)=$	0.50	L(T, black)=	0.33
black	12	12	24	0.50				
white	4	20	24	0.50	$Pr(C \mid black)=$	0.50	L(C, black) =	-0.33
	16	32	**48**					
priors	0.333	0.667						

23	triangle	circle		priors	$Pr(T \mid black)=$	0.55	L(T, black)=	-0.40
black	12	10	22	0.48				
white	22	2	24	0.52	$Pr(C \mid black)=$	0.45	L(C, black) =	0.40
	34	12	**46**					
priors	0.739	0.261						

24	triangle	circle		priors	$Pr(T \mid black)=$	0.55	L(T, black)=	0.38
black	12	10	22	0.48				
white	4	20	24	0.52	$Pr(C \mid black)=$	0.45	L(C, black) =	-0.38
	16	30	**46**					
priors	0.348	0.652						

25	circle	triangle		priors	$Pr(C \mid black)=$	0.55	L(C, black) =	-0.40
black	12	10	22	0.48				
white	22	2	24	0.52	$Pr(T \mid black)=$	0.45	L(T, black)=	0.40
	34	12	**46**					
priors	0.739	0.261						

26	circle	triangle		priors	$Pr(C \mid black)=$	0.55	L(C, black) =	0.38
black	12	10	22	0.48				
white	4	20	24	0.52	$Pr(T \mid black)=$	0.45	L(T, black)=	-0.38
	16	30	**46**					
priors	0.348	0.652						

27	triangle	circle		priors	$Pr(T \mid black)=$	0.60	L(T, black)=	-0.41
black	12	8	20	0.43				
white	24	2	26	0.57	$Pr(C \mid black)=$	0.40	L(C, black) =	0.41
	36	10	**46**					
priors	0.783	0.217						

28	triangle	circle		priors	$Pr(T \mid black)=$	0.60	L(T, black)=	0.40
black	12	8	20	0.43				
white	6	20	26	0.57	$Pr(C \mid black)=$	0.40	L(C, black) =	-0.40
	18	28	**46**					
priors	0.391	0.609						

29	circle	triangle		priors	$Pr(C \mid black)=$	0.60	L(C, black) =	-0.41
black	12	8	20	0.43				
white	24	2	26	0.57	$Pr(T \mid black)=$	0.40	L(T, black)=	0.41
	36	10	**46**					
priors	0.783	0.217						

30	circle	triangle		priors	$Pr(C \mid black)=$	0.60	L(C, black) =	0.40
black	12	8	20	0.43				
white	6	20	26	0.57	$Pr(T \mid black)=$	0.40	L(T, black)=	-0.40
	18	28	**46**					
priors	0.391	0.609						

Figure 6. Sets 31 to 40. Evidence: color (white); hypothesis: shape

31	triangle	circle		priors	Pr(T|white)=	0.50	L(T, white)=	-0.33
white	12	12	24	0.50				
black	20	4	24	0.50	Pr(C|white)=	0.50	L(C, white) =	0.33
	32	16	**48**					
priors	0.667	0.333						

32	triangle	circle		priors	Pr(T|white)=	0.50	L(T, white)=	0.33
white	12	12	24	0.50				
black	4	20	24	0.50	Pr(C|white)=	0.50	L(C, white) =	-0.33
	16	32	**48**					
priors	0.333	0.667						

33	triangle	circle		priors	Pr(T|white)=	0.55	L(T, white)=	-0.40
white	12	10	22	0.48				
black	22	2	24	0.52	Pr(C|white)=	0.45	L(C, white) =	0.40
	34	12	**46**					
priors	0.739	0.261						

34	triangle	circle		priors	Pr(T|white)=	0.55	L(T, white)=	0.38
white	12	10	22	0.48				
black	4	20	24	0.52	Pr(C|white)=	0.45	L(C, white) =	-0.38
	16	30	**46**					
priors	0.348	0.652						

35	circle	triangle		priors	Pr(C|white)=	0.55	L(C, white) =	-0.40
white	12	10	22	0.48				
black	22	2	24	0.52	Pr(T|white)=	0.45	L(T, white)=	0.40
	34	12	**46**					
priors	0.739	0.261						

36	circle	triangle		priors	Pr(C|white)=	0.55	L(C, white) =	0.38
white	12	10	22	0.48				
black	4	20	24	0.52	Pr(T|white)=	0.45	L(T, white)=	-0.38
	16	30	**46**					
priors	0.348	0.652						

37	triangle	circle		priors	Pr(T|white)=	0.60	L(T, white)=	-0.41
white	12	8	20	0.43				
black	24	2	26	0.57	Pr(C|white)=	0.40	L(C, white) =	0.41
	36	10	**46**					
priors	0.783	0.217						

38	triangle	circle		priors	Pr(T|white)=	0.60	L(T, white)=	0.40
white	12	8	20	0.43				
black	6	20	26	0.57	Pr(C|white)=	0.40	L(C, white) =	-0.40
	18	28	**46**					
priors	0.391	0.609						

39	circle	triangle		priors	Pr(C|white)=	0.60	L(C, white) =	-0.41
white	12	8	20	0.43				
black	24	2	26	0.57	Pr(T|white)=	0.40	L(T, white)=	0.41
	36	10	**46**					
priors	0.783	0.217						

40	circle	triangle		priors	Pr(C|white)=	0.60	L(C, white) =	0.40
white	12	8	20	0.43				
black	6	20	26	0.57	Pr(T|white)=	0.40	L(T, white)=	-0.40
	18	28	**46**					
priors	0.391	0.609						