Acta Psychologica 134 (2010) 162–174

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Preferences for different questions when testing hypotheses in an abstract task: Positivity does play a role, asymmetry does not

Paolo Cherubini [a,*], Patrice Rusconi [a], Selena Russo [b], Selenia Di Bari [a], Simona Sacchi [a]

[a] Department of Psychology, University of Milano-Bicocca, Italy
[b] Department of Cognitive Sciences and Education, University of Trento, Italy

ARTICLE INFO

Article history:
Received 15 August 2008
Received in revised form 8 January 2010
Accepted 22 January 2010
Available online 12 March 2010

PsycINFO classification:
2340

Keywords:
Hypothesis-testing
Information selection
Probabilistic reasoning

ABSTRACT

Previous studies on hypothesis-testing behaviour have reported systematic preferences for posing positive questions (i.e., inquiries about features that are consistent with the truth of the hypothesis) and different types of asymmetric questions (i.e., questions where the hypothesis confirming and the hypothesis disconfirming responses have different evidential strength). Both tendencies can contribute – in some circumstances – to confirmation biases (i.e., the improper acceptance or maintenance of an incorrect hypothesis). The empirical support for asymmetric testing is, however, scarce and partly contradictory, and the relative strength of positive testing and asymmetric testing has not been empirically compared. In four studies where subjects were asked to select (Experiment 1) or evaluate (Experiments 2–4) questions for controlling an abstract hypothesis, we orthogonally balanced the positivity/negativity of questions by their symmetry/asymmetry (Experiments 1–3), or by the type of asymmetry (confirmatory vs disconfirmatory; Experiment 4). In all Experiments participants strongly preferred positive to negative questions. Their choices were on the other hand mostly unaffected by symmetry and asymmetry in general, or – more specifically – by different types of asymmetry. Other results indicated that participants were sensitive to the diagnosticity of the questions (Experiments 1–3), and that they preferred testing features with a high probability under the focal hypothesis (Experiment 4). In the discussion we argue that recourse to asymmetric testing – observed in some previous studies using more contextualized problems – probably depends on context-related motivations and prior knowledge. In abstract tasks, where that knowledge is not available, more simple strategies – such as positive testing – are prevalent.

1. Introduction

Whenever we explicitly test the plausibility of a hypothesis, we ask questions either to ourselves, or to other people and external data bases. Since gathering all the evidence needed for an exhaustive check is seldom feasible, giving priority to some questions implies giving priority to some pieces of information above others. Different studies have emphasized that some human trends in gathering information might – in certain environments – cause undesirable side effects such as confirmation biases (Klayman & Ha, 1987; Nickerson, 1996, 1998; Wason, 1960, 1968) or the preservation of social stereotypes (Cameron & Trope, 2004; Trope & Thompson, 1997). The main goal of this study is to further explore two of those tendencies, namely the alleged preference for posing *asymmetrical questions* and *positive questions* (the latter, also known as "congruent" questions, Baron, Beattie, & Hershey, 1988), in order to verify their actual occurrence and to compare their relative strengths in abstract tasks, where domain-specific motivations and prior knowledge are hardly accessible.

2. Positivity and asymmetry of questions

A common definition describes a positive question as a question where a positive response ("yes") supports the truth of the hypothesis (Klayman, 1995; Klayman & Ha, 1987; Snyder & Swann, 1978). Posing positive questions does not however necessarily imply an ability to anticipate the epistemic effects of the "yes" or "no" responses. They can more simply originate from a tendency to inquire about features that "match" the hypothesis, i.e., features that are more typical of instances where the hypothesis is true, than of instances where it is false. When investigating whether a target individual is an extrovert, for example, asking "does she like parties?" is a positive question. The inquired feature matches the representation of an extrovert, and – as a result – a "yes" supports the hypothesis of extroversion, while a "no" weakens it. By contrast,

* Corresponding author. Address: Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milano, Italy. Tel.: +39 02 6448 3811; fax: +39 02 6448 3706.
E-mail address: paolo.cherubini@unimib.it (P. Cherubini).

0001-6918/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2010.01.007

asking "does she enjoy long solitary walks?" is a negative question: the feature matches the representation of an introvert, and accordingly a "yes" weakens the hypothesis of extroversion, and a "no" supports it. Symmetry/asymmetry of questions is more complex, as it is a matter of the quantity of information received and not only of its valence. An asymmetric query can – depending on the answer it receives – confirm a hypothesis *more* than it can disconfirm it (asymmetrically *confirming* questions; or "high risk" testing strategies, Poletiek & Berndsen, 2000), or vice versa (asymmetrically *disconfirming* questions; also known as "extreme" tests, Skov & Sherman, 1986; Slowiaczek, Klayman, Sherman, & Skov, 1992; or "low risk" testing strategies, Poletiek & Berndsen, 2000). Investigating, for instance, the extroversion of a person by asking "is she always the life of parties?" is an asymmetrically confirming test: a "yes" response is improbable, but, if received, would strongly support the hypothesis. On the other hand, a "no" response is probable, but – if received – only weakly disconfirms the hypothesis. Similarly the question "does she love spending most Saturday evenings reading poetry by herself?" is asymmetrically disconfirming (a "yes" strongly falsifies extroversion, whereas a "no" only weakly confirms it). For a stricter definition of asymmetry, many quantitative measures of the strength of confirmation are available (Crupi, Tentori, & Gonzalez, 2007). The first, most common, and easiest one is the Bayes' factor. A common formal description of logically sound belief revision is the Bayes' rule. It states how a degree of belief in the truth of a hypothesis H (expressed as the epistemic probability that H is true) should be revised in light of a body of newly acquired evidence. A useful formulation of the Bayes' rule is in terms of odds and likelihood ratios (Beyth-Marom & Fischhoff, 1983; Fischhoff & Beyth-Marom, 1983; Good, 1950, 1960, 1983; Osteyee & Good, 1974):

$$\frac{p(H|E)}{p(\neg H|E)} = \frac{p(H)}{p(\neg H)} \times \frac{p(E|H)}{p(E|\neg H)}$$

where $p()$ is read "probability of", H means "the hypothesis is true", $\neg H$ means "the hypothesis is false", E is the body of evidence, and the | symbol stands for a conditional probability (it can be read "given"). Reading from the left, the three terms of the formula are:

(a) the posterior odds: the ratio between the probability that H is true given E and the probability that H is false given E;
(b) the prior odds: the ratio between the probability that H was true before acquiring E, and the probability that it was false;
(c) the Bayes factor, that is the likelihood ratio of E (thereafter, LR): the ratio between the probability of observing E assuming the truth of H and the probability of nevertheless observing E if H were false.

Bayes' rule is a straightforward and undisputed consequence of the basic axioms of standard probability calculus. More importantly, the LR appropriately describes an intuition that is common in many fields where correctly weighing evidence is critically important, such as medical diagnosis or legal judgement. A piece of evidence (e.g., a symptom, a clue) that is equally probable regardless of whether H (e.g., a possible diagnosis, a charge of wrongdoing) is true or false, does not change the probability that H is true or false, and therefore it is uninformative. Such a piece of evidence, with $p(E|H) = p(E|\neg H)$, has LR = 1, and thus leaves the posterior odds unchanged with respect to the prior odds. Along the same lines, a piece of evidence with LR > 1 increases the posterior probability of H with respect to $\neg H$: it is thus *confirmatory*. Finally, a piece of evidence with LR < 1 decreases the posterior probability of H with respect to $\neg H$: it is *disconfirmatory*. A dichotomous question – namely one accepting only "yes/no" as mutually exclusive answers – is symmetric if and only if the two answers have the same LR (a "yes" confirms H exactly as much as a "no" confirms $\neg H$, or vice versa). Otherwise, it is asymmetric.

Symmetric queries are "fair" questions, with equal chances[1] of either confirming or disconfirming the hypothesis by the same amount of evidential strength. By choosing them, inquirers do not commit themselves either to a conservative or to a non-conservative stance (in technical terms, they equate the risk of incurring in a Type II, false negative, or a Type I, false positive, error). Asymmetric confirming tests have a relatively low probability of yielding strong evidence in support of the hypothesis, and a correspondingly high probability of finding weak evidence that refutes it. They are conservative questions, as they maximize the chances of (weakly) rejecting the hypothesis, while minimizing those of (strongly) accepting it. Asymmetric confirming queries shift the balance in favour of Type II, false negative errors, and accordingly should be typical of contexts where there are good reasons to prefer Type II errors to Type I errors (e.g., when evaluating a crime charge in a judicial setting). By contrast, asymmetric disconfirming questions have a relatively low probability of finding strong evidence that disconfirms the hypothesis, and a correspondingly high probability of yielding weak evidence in support of it. By making probable a weak confirmation at the expense of an improbable strong refutation of the hypothesis, they denote a preference for risking Type I instead of Type II errors: a typical attitude of some preliminary medical screening tests (such as the PSA test for prostate cancer), or of the "overprotecting" policy in antiterrorism airport checks (Hammond, 2007). An alternative – and common – name for these latter sort of questions is "extreme tests" (Skov & Sherman, 1986), meaning that they address the feature that has the most extreme probability (either high or low) under the focal hypothesis, and the less extreme probability under the alternative one.

The properties of symmetric or asymmetric queries are formally independent of their positivity/negativity: that is, a question can be symmetric, asymmetric confirming, or asymmetric disconfirming, disregarding whether the response that supports the hypothesis is "yes" or "no".

In theory, the best questions to posit are the most *diagnostic* ones, those with a maximal expected utility in informational terms. It can be measured as the mean LR – that is the weighed average of the LR of the confirming response in support of the hypothesis and the LR of the disconfirming response in support of the alternative hypothesis[2]:

$$\text{Mean LR} = p(\text{confirmation}) * \text{LR}(\text{confirmation}) + p(\text{disconfirmation}) * \frac{1}{\text{LR}(\text{disconfirmation})}$$

By not choosing the most diagnostic questions among the available ones, a person risks to throw away useful information, thus increasing the chances of avoidable errors. Diagnosticity is never affected by the positivity/negativity of the query. Furthermore, it is not systematically affected by its symmetry/asymmetry: depending on the parameters associated to the tested features, there can be symmetric and asymmetric questions of equal diagnosticity,

[1] Whilst the properties of LRs are independent of the priors $p(H)$ and $p(\neg H)$, the probabilities of receiving either a confirming or a disconfirming answer depend upon $p(H)$. Where we discuss them in this paper we assume that $p(H) = 0.5$, a common assumption in most other previous studies on this topic, and a premise in our experiments.

[2] Mean LR – or the mean weight of evidence (WE; see Appendix C) – is a rough measure of the overall utility of a question, even though it is widely used in literature (e.g., Slowiaczek et al., 1992; Trope & Bassok, 1982). A more proper measure is the less intuitive expected information gain (IG), computed as a difference in informational entropy before and after having received an answer to the question (Oaksford & Chater, 2007). In this case, as in most others where the prior probabilities of the hypotheses are the same, overall usefulness as measured by expected IG (or any other measure that has been proposed, such as Kullback-Leibler's numbers) is directly proportional to the usefulness as measured by mean LR – that therefore is a viable measure.