INCONSISTENT PROBABILITY ESTIMATES OF A HYPOTHESIS: THE ROLE OF CONTRASTING SUPPORT

Nicolao Bonini
Department of Economics and Cognitive Science Laboratory, University of Trento

Michel Gonzalez
Laboratoire de Psychologie Cognitive, URA CNRS 182
Université de Provence

Running Head: Inconsistent probability estimates

Corresponding author:

Nicolao Bonini
Laboratorio di Scienze Cognitive
Università di Trento
Via Tartarotti, 7
38068 Rovereto, Italy
nbonini@form.unitn.it
Phone: (++ 39)0464-48.3525
Fax: (++ 39)0464-48.3514

INCONSISTENT PROBABILITY ESTIMATES OF A HYPOTHESIS: THE ROLE OF CONTRASTING SUPPORT

Abstract

This paper studies consistency in the judged probability of a target hypothesis in incomplete lists of mutually exclusive hypotheses. Specifically, it controls the role played by the support of displayed competing hypotheses and the relatedness between the target hypothesis and its alternatives. Two experiments are reported. In both, groups of subjects were presented with a list of mutually exclusive non-exhaustive causes of a person's death. For each cause, they were asked to judge the probability rate of causing the person's decease. Target causes were presented in all lists. Several other alternative causes to the target ones differed across the lists. Findings show that the judged probability of a target cause changes as a function of the support of the displayed competing causes. Specifically, it is higher when its competing displayed causes have low rather than high support. Findings are consistent with Support Theory (Tversky and Koehler, 1994; Rottenstreich and Tversky, 1997).

Key words: probability judgement; extensionality; consistency; salience; availability; support theory.

In various situations, when several alternative hypotheses can explain a given evidential fact (e.g., a car failing to start), a probability evaluation of each of them is made. In most situations, a person who considers an evidential fact retrieves causal hypotheses from memory. In most studies of probability judgments, however, a number of specific hypotheses are selected from the set of possible hypotheses, and displayed for evaluation (Fischhoff, Slovic and Lichtenstein, 1978; Harries and Harvey, 2000; Russo and Kolzow, 1994; Tverky and Koehler, 1994; Van Schie and Van der Pligt, 1990).

Studies of probability judgments of a set of causal hypotheses have shown that the judged probability of a target hypothesis depends on the set of displayed hypotheses (Bonini, Van der Pligt, Van Schie and Legrenzi, 1994; Fischhoff, Slovic and Lichtenstein, 1978; Van der Pligt, Eiser and Spears, 1987; Van Schie and Van der Pligt, 1990). For example, Fischhoff et al. asked car mechanics as well as lay people to predict the relative frequencies of a number of possible causes for a car failing to start. The same cause (e.g., “battery”) was displayed in two lists that covered a selection of three or six specific causes of the failure. Its frequency was judged higher when fewer causes were displayed.

Such a phenomenon is embarrassing because it casts doubts on the reliability of probability judgment. Indeed, for the same evidential fact, the probability of a hypothesis should not depend on the set of displayed hypotheses. There is no reason to believe that the “battery” cause is a more or less probable cause of the car failing to start as a function of which specific alternatives are selected.

The aim of the paper is twofold. First, we document inconsistent probability judgments of a target hypothesis presented in lists of non-exhaustive and mutually exclusive hypotheses. Second, we provide an interpretation for such a phenomenon.

How could the results in literature, which have shown inconsistent probability estimates of a target hypothesis as a function of the set of competing hypotheses, be interpreted? Below, we outline a few explanations by grouping them into three main categories. The first two relate to several interpretations that are given in literature (mainly in fault tree literature), and that can be used to account for the inconsistency phenomenon: the first refers to a specific experimental constraint, the second to the representation of the target hypothesis. The third category relates to the judged support to the target and competing hypotheses.

The side effect of the experimental demands. According to this interpretation, inconsistency is a consequence of a constraint imposed by the experimenter. In fault trees and related studies, people are asked to make sure that their probability or frequency estimates sum up to 100%. This constraint is justified by the additivity requirement because people are presented with a list of exhaustive and mutually exclusive causes[(1).]

For example, in Fischhoff et al. seminal work, both in the full and pruned trees participants were presented with an exhaustive list of possible causes for a car failing to start. Specifically, in the full fault tree six specific causes were displayed ("battery", "fuel system", "engine", "starting system", "ignition system", "mischief"), plus the catch-all category "other causes". In the pruned fault tree, only three specific causes from the previous tree were presented, plus the catch-all category. Two main findings were reported. First, participants underestimated the probability of the catch-all category in the pruned tree. From the additivity of probability it follows that the probability judgment of the catch-all category in the pruned tree should be equal to the sum of the probabilities of the omitted causes plus the probability of the catch-all category in the full tree. However, Fischhoff et al. found that the probability of the catch-all category in the pruned tree was lower than normatively expected (.22 vs. 45). Second, and most relevant for our paper, they found inconsistent probability judgments for specific causes displayed in both trees: the sum of the probability estimates of the three target causes increased from the full (.55) to the pruned tree (.78).

Although fault tree and subsequent related studies were focused on the catch-all underestimation bias, the underestimation of the catch-all category in the pruned tree will necessarily determine an equivalent overestimation of the target hypotheses in the same tree because of the sum-to-100% probability requirement. Thus, according to this interpretation, the inconsistent probability judgments of a hypothesis found in fault tree and related studies are a side effect of the catch-all underestimation bias and the sum-to-100% probability constraint. It should be noted that people exhibit non-additive probability distributions (Teigen, 1983; Tversky and Koehler, 1994; Wright and Whaley, 1983). It is thus questionable whether the inconsistency phenomenon first reported by Fischhoff et al. could be found in the absence of such a constraint.

Another interpretation has been advanced to explain both the catch-all underestimation bias and the inconsistent judgments of a hypothesis in fault tree and

related studies (Van der Pligt, Eiser and Spears, 1987; Van der Pligt, Van Schie and Hoevenagel, 1998). This interpretation is based on the anchoring and adjustment heuristic (Kahneman and Tversky, 1973), and again requires the 100% probability constraint to be applied. To illustrate it, let us consider the following. When assessing a cause on a list of exhaustive and mutually exclusive causes, people start with an anchor-value that is later adjusted. The anchor-value is approximately equal to 100% divided by the total number of displayed causes. As a consequence, the fewer the causes displayed in the tree the higher the anchor-value. Because of an insufficient adjustment of the anchor-value, the use of this heuristic would explain why the probability estimates of a hypothesis (including the “catch-all” category) increase from the full to the pruned tree.

Hypothesis redefinition. This explanation does not rely on the sum-to-100% probability constraint. In this interpretation, the inconsistency phenomenon is due to a change in the representation of the target hypothesis across lists or trees.

Hirt and Castellan (1988) argued that omissions in a fault tree could induce a subject to redefine the content of a target hypothesis. For example, the cause "battery" could be redefined as a more extensive category in the pruned than in the full tree by covering instances of the omitted cause "starting system". This would explain the inconsistency in the probability estimates of the target hypothesis across the trees as well as the catch-all underestimation bias.

Another explanation of inconsistency in probability judgments assumes that people spontaneously organise the displayed hypotheses before assessing them. In this case, the failure of consistency would not be due to a change in the content of the hypothesis but rather to a category organisation. For example, when several instances of a category of causes of death are displayed (e.g., several types of cancer), people may well group them into the same category (e.g., "cancer"), and later assess the inclusive category rather than each single instance.

The Similarity Coverage Model (Smith, Shafir and Osherson, 1993) posits a similar category-generation-and-evaluation process. According to this model, when a subject assesses the probability that a conclusion is true given that a premise is true, she generates and assesses the lowest-level category that includes the premise and conclusion categories. For example, when a subject is asked to judge the probability that "bobcats use serotonin as a neurotransmitter given that tigers and cougars do" she

will assess the probability that the inclusive category (e.g., "feline") will use serotonin as a neurotransmitter. Task restructuring operations are frequent in probability judgment and choice tasks. These editing operations are believed to help the subject simplify her judgments or choices (Coupey, 1994; Henderson and Peterson, 1992; Kahneman and Tversky, 1979).

Bonini, Van der Pligt, Van Schie and Legrenzi (1994) reported evidence of inconsistent probability judgments of a target hypothesis that could be due to a spontaneous grouping of displayed causes. In that study, the probability judgment of the target hypothesis “lung cancer” was lower when it was displayed together with three other examples of causes of death by cancer (multiple cancer list) than when it was the only cause of death by cancer (single cancer list). This finding could be explained by saying that subjects in the multiple cancer list grouped the four specific types of cancer into the inclusive category "cancer", assigned a probability value to this category and distributed the probability value across its four types. As a consequence of this categorisation-and-probability-redistribution process, the probability of "lung cancer" could well be lower in the multiple than in the single cancer list.

Focal and contrasting support. We propose an interpretation of the inconsistency of probability judgements of a target hypothesis presented in different sets of alternative hypotheses. This interpretation is conditioned neither by the specific features of the judgment task, such as the requirement of the sum-to-100% of the probability estimates, nor by the different number of hypotheses displayed in the sets. Nor is it based on a different representation of the target hypothesis in the sets. This interpretation lies on Support Theory, and is based on the notion of “contrasting support” (cf. Rottenstreich and Tversky, 1997; Tversky and Koehler, 1994).

According to Support Theory, the probability judgment of a focal hypothesis H is based on its degree of support (focal support) and the degree of support of its alternative not-H (contrasting support). Focal support positively contributes to the judged probability of the focal hypothesis, whereas its contrasting support negatively contributes to the judged probability[(2)]. For example, people may be asked to judge the probability of dying by “lung cancer rather than thrombosis”. According to Support Theory, the judged probability of the focal hypothesis “lung cancer” equals the ratio between focal support and total support (e.g. the sum of support for “lung cancer” and “thrombosis”).

Not-H is often badly defined in terms of an exhaustive list of specific alternative hypotheses to H, such as in fault trees and related studies. In these situations, the list of specific hypotheses is incomplete, and it is ended by a catch-all category such as “other causes”. In these situations, the representation of not-H depends on the retrieval of facts from memory in order to characterise the implicit catch-all category, and on the alternative hypotheses that have already been displayed - for example, as when the suggested competing hypotheses could never have spontaneously been thought of. Thus, the display of specific competing hypotheses contributes to the representation of not-H, and, as a consequence, to its judged support. If the displayed alternatives have low support, we assume that the support for not-H will be low. By contrast, if the support for the displayed alternative is high, the support for not-H will be high.

We suppose that the focal hypothesis H is well defined in different sets of competing hypotheses, given that it is characterised by the same label. Therefore, its focal support should not depend on the specific set of displayed alternatives.

To illustrate this, let us consider a situation where "lung cancer" is judged in a list covering "infarctus" and "road accident" as two examples of a number of possible competing hypotheses with a high degree of support. The same cause “lung cancer” could however be presented in a list covering “influenza” and “diabetes” as two examples of a number of possible competing hypotheses with a low degree of support. We expect that the support for "lung cancer" does not change across the two lists, but that the support for the competing hypotheses to “lung cancer” does (e.g., “other than lung cancer”). Specifically, we expect it to be higher in the former than in the latter list.

Following Support Theory, we predict that the judged probability of “lung cancer” will be lower in lists displaying high rather than low competing support hypotheses. Inconsistent probability judgments of a focal hypothesis can thus be attributed to differences in its contrasting support.

The paper is organised as follows. In Experiment 1, the effect of the factors "display of contrasting support" and "relatedness between presented hypotheses" is tested. In order to exclude other potential effects, the same number of hypotheses is displayed across conditions, and the sum-to-100% probability requirement is absent. Moreover, the mix of general and specific hypotheses is eliminated in order to avoid category redefinition due to vertical incongruity between displayed hypotheses (see Russo and Kolzow, 1994 for a discussion of this kind of restructuring activity). In Experiment 2,

the role played by the display of contrasting support in the consistency of probability judgments of a target hypothesis is further controlled.

EXPERIMENT 1

According to the contrasting support interpretation, the judged probability of a target hypothesis is higher when low rather than high competing support hypotheses are displayed. According to the relatedness-between-hypotheses interpretation, the judged probability of a target hypothesis is lower when it and its competing hypotheses can be easily grouped together than when they cannot. In the former situations, people first organise the displayed hypotheses within the same inclusive category, then assess the inclusive category, and finally distribute the global value across its instances.

In order to distinguish the effects of the contrasting support factor and the relatedness-between-hypotheses factor, three target hypotheses were presented in different lists. In each list, competing hypotheses had high or low support, and belonged / did not belong to the same category as the target hypothesis. According to the contrasting support interpretation, it is predicted that all three target hypotheses will be affected by the degree of support of the displayed competing hypotheses, regardless of the presence or absence of relatedness between them. On the contrary, according to the relatedness interpretation, it is predicted that only the judged probability of the target cause that belongs to the same category as the competing causes will be affected.

Method

Following Tversky and Koehler (1994, Study 1), students were asked to assess the likelihood of various possible causes of death of a single person. Four lists of causes of death were prepared and one list was presented to one of four groups of participants. Each list displayed six mutually exclusive non-exhaustive possible causes of death of a person. Participants judged the percentage probability of each cause being the cause of death of an Italian man aged 40-45. Three causes were inserted in all the lists (target hypotheses). The other three causes of each list were distinct in four different sets (competing hypotheses). They described either three specific types of cancer or three different kinds of disease. All of them had either a low or a high degree of support as a possible cause of death.

Stimulus materials

A pilot study was run to devise the four lists of possible causes of death. A group of 50 undergraduate students of the University of Cagliari (Italy), who did not participate in the experiment, were presented with a list of 46 items. These items were selected from fault tree and related studies about a person's decease (Bonini et al., 1994; Russo and Kolzow, 1994). Each item described a possible cause of death and was organised across three categories of causes of death: 14 items described different types of cancer; 24 items described other types of disease; 8 items described accidental causes of death. Participants were asked to estimate the percentage of death of Italian men aged 40-45 by each cause.

Findings from the pilot study show that the total judged frequency largely exceeded 100% (mean total frequency: 418%).

Selection of the three target causes. The causes of death that were judged as most frequent in each of the three categories were "lung cancer" (mean frequency judgment: 27%), "infarctus" (28%), and "road accident" (28%). These three items were consequently selected as target causes. That is, they were inserted in each of the four lists that were used in the experiment.

Selection of the competing causes. Four other sets of three items each were selected. One set contained the causes of death by cancer that were judged as most frequent after "lung cancer": "leukaemia" (19%), "liver cancer" (16%), and "stomach cancer" (15%). Another set contained the causes of death by cancer that were judged as least frequent: "skin cancer" (10%), "oesophagus cancer" (9%), and "larynx cancer" (9%). A third set contained the causes of death by a disease other than cancer that were judged as most frequent after "infarctus": "AIDS" (28%), "cerebral ischemia" (13%), and "cirrhosis" (10%). The last set contained the causes of death by a disease other than cancer that were judged as least frequent: "asthma" (2%), "abdominal hernia" (1%) and "influenza" (1%). Only one of these four sets of competing causes was inserted in each of the four lists.

In sum, each list contained three target causes and three competing causes. The three competing causes were either types of cancer or kinds of disease other than cancer that were judged as either frequent or rare causes of death.

Participants

200 undergraduate students of the University of Cagliari (Italy) took part in the experiment. They were randomly assigned to four equal-sized groups (N = 50 per group).

Design and Procedure

The four lists correspond to the crossing of two factors: Support of displayed competing causes (henceforth "Competing Support"; high vs. low support of displayed competing causes) and Type of competing causes (henceforth "Type"; types of cancer vs. diseases other than cancer). A 2 (competing support) x 2 (competing type) x 3 (target cause) experimental design was used, with the last factor as a within-subject factor. Each group was presented with only one list with the six items printed on one page, in one of two different orders of presentation. Participants read the following instructions:

> Last year thousands of Italian men aged 40-45 died of various causes of death. Imagine that one of these men be randomly selected. In the next page, you will be presented with six of the many possible causes of his death.
>
> You are asked to estimate the probability that, in your opinion, the decease of that person was due to each of the six causes. For each cause, write a probability estimate between 0% (you think that the decease is surely not due to that cause) and 100% (you think that the decease is surely due to that cause).
>
> According to statistics, last year, for example, the probability of death by “suicide” and “inflammation” in the Italian population was 1.5% and 4.2% respectively.

Results

In the four considered conditions, the mean sum of the probability judgments of the six non-exhaustive causes was equal to 1.31. Hence, probabilities were overestimated.

The mean probability estimates of the three target causes, and the total probability estimates of the competing causes across the four conditions are presented in Exhibit 1.

As expected, the judged probability of the competing causes was higher in the high than in the low support condition. Moreover, in the low competing support condition the judged probability of competing types of cancer was higher than competing non-cancer diseases ($F(1, 196) = 3.04$, $p = .08$).

An ANOVA on a 2 (Competing support) x 2 (Competing type) x 3 (Target cause) design with the last factor as a within-subject factor was performed.

Exhibit 1 about here

As predicted from the contrasting support hypothesis, the mean probability judgments of the target causes were higher in the low than in the high competing support condition (.36 vs. 30, respectively; $F(1, 196) = 6.91$, $p < .01$). However, the effect of Competing Support interacts with the Type of Competing Causes ($F(1, 196) = 8.14$, $p < .005$). Competing Support affects probability judgment consistency when competing causes are non-cancer diseases (.41 vs. .28 for low and high competing support, respectively; $F(1, 196) = 15.02$, $p < .0001$), but it does not when competing causes are types of cancer (.30 vs. .31; $F(1, 196) < 1$). No statistical significant interaction between Competing Support and Target cause was found ($F(2, 392) = 2.35$, $p < .10$). This finding suggests that the effect of the Competing Support factor concerns all target causes (see Exhibit 1).

As predicted by the relatedness-between-hypotheses interpretation, the judged probability of "lung cancer" is lower when the list displays competing types of cancer than in the case of competing non-cancer diseases (see Exhibit 1). However, the difference is weak (.30 vs. .35, respectively; $F(1, 196) = 3.68$, $p = .06$). Moreover, a significant interaction between Type of competing causes and Competing Support on the judgments of "lung cancer" is found ($F(1, 196) = 5.66$, $p < .05$). This shows that the relatedness-between-hypotheses effect depends on the support for the displayed competing causes. The judged probability of "lung cancer" is lower when the competing causes are types of cancer (.30) than when they are non-cancer diseases (.43), but only in the low competing support condition ($F(1, 196) = 9.24$, $p < .005$). Furthermore, as shown in the left part of Exhibit 1, in this condition the effect of Type of competing causes is not limited to "lung cancer" but it concerns all the target causes ($F(1, 196) = 10.43$, $p < .001$), and there is no significant interaction between Type and Target ($F(1, 196) < 1$). In sum, the effect of the Type factor of competing causes (cancer vs. non-cancer causes of death) cannot be interpreted as due to the relatedness between the target and the displayed competing causes.

Discussion

Results of Experiment 1 show that the judged probability of a target hypothesis changes as a function of the support of the displayed competing hypotheses. This effect is rather different from the one evidenced in fault tree studies because in our experiment the number of displayed hypotheses is the same across the lists, and the sum-to-100% probability constraint is absent. Two explanations were considered to account for this: contrasting support and relatedness-between-hypotheses.

From the relatedness-between-hypotheses based explanation, the judged probability of a target hypothesis is lower when it and its competing hypotheses can be grouped into the same inclusive category. However, this explanation does not give an adequate account of our results. First, the effect of the relatedness-between-hypotheses on the consistency of probability judgements of "lung cancer" should have been found regardless of the support of the displayed competing causes. Unlike this prediction, it was found only in the low competing support condition. Second, the relatedness effect should have been found only for "lung cancer”. Again, unlike this prediction, it was found for the three target causes .

According to the contrasting support explanation, the judged probability of a target hypothesis is higher when the displayed competing hypotheses have low rather than high support. This prediction was confirmed when the displayed competing hypotheses were non-cancer diseases. No effect was found when the displayed competing hypotheses were various kinds of cancer.

As a contribution to the discussion, we propose two possible interpretations for the lack of effect of Competing Support when the displayed competing hypotheses are various kinds of cancer. These two interpretations are compatible with the contrasting support hypothesis.

Let us consider a target cause such as "road accident". In both the high and the low competing support conditions where the displayed competing causes were types of cancer, four kinds of cancer and one other disease ("infarctus") competed with "road accident". In both competing support conditions, the names of the four types of cancer might easily evoke the inclusive category "cancer". Thus, in both competing support conditions "road accident" might have been compared to "cancer" and “infarctus” as alternative causes of death. In this case, the contrasting support for "road accident" depends on the support for "cancer" rather than for the displayed specific types of

cancer. As a consequence, the contrasting support for "road accident" is the same across the two competing support conditions, and the probability judgment of "road accident" (or of "infarctus") should not vary.

The other interpretation does not assume a category organisation process, but is based on the weakness of the experimental manipulation of competing support. Remember that three target causes with a high degree of support were presented in all conditions. Let us consider, for example, the target cause “infarctus”. In the high competing support condition, the displayed alternatives to “infarctus” were the other two target causes and the three competing causes with high support. Thus, “infarctus” was presented with five alternatives that all had high support. However, in the low competing support condition the displayed alternatives to “infarctus” were the same two target causes and the three competing causes with low support. Thus, in the low competing support condition not all the displayed causes had low support. As a consequence, the difference in the total competing support for “infarctus” - and for the other target causes alike - between the two conditions is weak. Thus, the lack of substantial difference in total support across the two competing support conditions may account for why the effect was not always found in this experiment.

EXPERIMENT 2

Experiment 1 has shown that relatedness-between-hypotheses had not a noticeable effect on probability judgment consistency, and could not systematically account for the reported findings. Thus, the relatedness hypothesis is no longer considered in this experiment.

In Experiment 1, an effect of contrasting support was evidenced, but it was not found in all conditions. The aim of this experiment was to better control the contrastive support hypothesis. To this end, two lists were devised in order to maximise the difference in the total contrastive support for the target hypotheses across conditions, and to avoid the grouping of displayed hypotheses. Unlike the previous experiment, only two target causes were used: one with high and one with low support. Thus, in this experiment the effect of contrasting support was tested on target causes of different degrees of support.

Method

Two lists of five causes of death were presented to two groups of participants. Each list displayed five mutually exclusive non-exhaustive possible causes of death of a person. As in the previous experiment, participants judged the percentage probability that each of the displayed items might be the cause of a person's decease.

Stimulus materials

The lists were based on the results of the pilot study that has been described in Experiment 1. Two items (target causes) were common to both lists: one with high support ("lung cancer") and the other with low support ("diabetes"). In the pilot study, the mean probability judgments of dying by "lung cancer" and "diabetes" were .27 and .05. In each list, the three remaining items were specific causes of death that did not belong to the same inclusive category of target causes. In the low competing support condition, the three specific items were "asthma", "abdominal hernia" and "influenza". They are the same causes as the ones used in the previous experiment, which had been judged unlikely in the pilot study (mean total judged probability: .04). In the high competing support condition, the three specific items were "AIDS", "road accident", and "infarctus", which were judged in the pilot study as the most likely causes of death together with "lung cancer" (mean total judged probability: .84).

Participants

The participants were 134 students from the University of Cagliari (Italy). Six more students had been discarded from the analysis because they had either not answered or assigned a 100% probability estimate to a hypothesis and a non-null probability to another hypothesis. The 134 students were randomly assigned to two groups of 64 (low competing support condition) and 70 (high competing support condition).

Design and Procedure

The experimental design was a 2 (competing support) x 2 (target cause) design, with the last factor as a within-subject factor.

The instructions and the probability estimation task were the same as in the previous experiment. In each condition, the group of participants was split in two subgroups, with the five items presented in two different orders.

Results

The mean sum of the probability judgments of the five causes was equal to 0.97. It was higher in the high than in the low competing support condition (1.16 vs. 0.76, respectively; $F(1, 132) = 24.17$, $p < .001$).

The mean probability estimates of the two target causes, and the total probability estimates of the competing causes across the conditions are presented in Exhibit 2. An ANOVA on a 2 (competing support) x 2 (target cause) design was performed, with the last factor as a within-subject factor.

Exhibit 2 about here

As expected, the judged probability of the target causes was higher in the low than in the high competing support condition ($F(1,132) = 31.49$, $p < .001$). As the two target causes have different levels of expectation, the competing support effect is less important for "diabetes" than for "lung cancer". Nonetheless, it is statistically significant in each case ($F(1,132) = 5.92$, $p < .02$, and $F(1,132) = 32.95$, $p < .001$, respectively).

GENERAL DISCUSSION

The reported findings show that in incomplete lists of hypotheses the judged probability of a target hypothesis changes as a function of the display of its alternatives. Specifically, it is higher when low rather than high support competing hypotheses are displayed.

These inconsistent probability judgments are normatively incorrect. Specifically, they demonstrate a lack of extensionality of subjective probability when people are asked to assess the cause of death of a person in lists that describe the person and the cause in the same way (e.g., the probability that a 40/45-year-old Italian man died by "lung cancer" rather than by any other possible cause of death).

The effect of the competing support display is accounted for by the contrasting support hypothesis that lies within Support Theory (Tversky and Koehler, 1994; Rottenstreich and Tversky, 1997).

As a contribution to the discussion, we argue that two mechanisms may be responsible for the differently judged contrastive support across the incomplete lists of hypotheses: a memory-based or a salience mechanism. In the first case, when low support competing hypotheses are displayed, people may fail to retrieve high support ones. In the second case, high support hypotheses may well be retrieved, but they are weighed less in the low than in the high competing support display because they were not explicitly presented. Tversky and Koehler (1994) also indicated these two mechanisms as sources of lack of extensionality in probability judgment.

Koehler (1994) provided further evidence of memory retrieval as a source of failure of invariance in the judged probability of a hypothesis. The reported findings show that the subjects who generated their own hypotheses expressed less confidence in their truth than the subjects who were presented with the same hypotheses for evaluation. The author argues that "[in the generation condition, subjects] should consider more alternative hypotheses than people who are merely asked to evaluate a hypothesis that is presented to them" (p. 462). This interpretation is consistent with the contrasting support notion. The findings reported by Koehler (1994) could be due to a difference in the judged contrasting support across the generated vs. evaluated condition because of differences in the retrieval of competing hypotheses across the same conditions. As a consequence, confidence judgments would be higher in the displayed than in the generated condition.

Finally, although our paper is not concerned with the accuracy of probability judgment, the distinction between metrics and mappings components of a real-word quantitative estimation (Brown and Siglier, 1993) is useful to discuss our results. According to Brown and Siglier, a "quantitative estimation is derived from two types of knowledge: knowledge of the distribution of quantitative values in the domain, and knowledge of the relative positions of particular entities within that distribution" (p. 529). Information about central tendency, variability, range, or shape of a distribution relates to metrics knowledge. Information about relative status of individual entities such as the rank order of specific causes of death relates to mapping knowledge.

The effect of the competing support display is not associated with a change in the rank order of the target causes, but with a change in the magnitude value of the same causes. The reported findings show that the rank order of a cause was stable across the two competing support displays (e.g., in both displays "lung cancer” was ranked before

"diabetes", and "road accidents" was ranked before "lung cancer" or "infarctus"). Furthermore, the ratios of the judged probabilities of the target causes were similar across the two displays (e.g., in both displays the judged probability of "lung cancer" was nearly 3 times higher than “diabetes").

These findings suggest that the focal support of a cause does not change across different displays of competing causes. For example, "lung cancer" has more/better reasons than "diabetes" to be a cause of death, regardless of which alternative causes it is compared to. Thus, these findings suggest that the mapping knowledge of a set of causes would be a function of focal support, whereas the final quantitative estimation (e.g., the amount of chances, the absolute or relative frequency) would be a function of focal and competing support. However, more research is needed to disentangle the role of these two types of support in the derivation of a real-world quantitative estimation.

FOOTNOTES

(1) A fault tree is a list that hierarchically organises several categories of causes related to an evidential fact such as a car failing to start, the decease of a person or a firm bankruptcy. The systematic analysis of the problem by a fault tree is believed to enable better and more reliable judgement. The categories of causes of the major event are usually named branches of the tree (for a review on fault tree studies, see Russo and Kolzow, 1994).

(2) A similar analysis of the focal probability judgment is made by Extended Support Theory (cf. Idsen, Krantz, Osherson and Bonini, 2001). In this theory, the judged probability of a focal proposition X is based on the perceived evidence in support of X and against X. However, unlike Support Theory, -Extended Support Theory allows for the case where evidence against X is not considered by the judge.

ACKNOWLEDGEMENTS

We wish to thank two anonymous referees for their comments on an earlier version of this paper. We also wish to thank Peter Ayton, Dave Krantz, Paolo Legrenzi, Dan Osherson and Yuval Rottenstreich for helpful discussions of this study with the first author. Finally, we gratefully acknowledge a financial contribution of MURST assigned to the first author.

REFERENCES

Bonini, N., Van Schie, E.C.M., Van der Pligt, J., Legrenzi, P. 1994. Gli effetti della manipolazione del contesto di giudizio sull'invarianza della stima probabilistica dell'accadimento di eventi. *Giornale Italiano di Psicologia*, 1, 47, 62.

Brown, N. R., Siegler, R. S. 1993. Metrics and Mappings: A Framework for Understanding Real-World Quantitative Estimation. *Psychological Review*, 100, 3, 511-534.

Coupey, E. 1994. Restructuring: Constructive Processing of Information Displays in Consumer Choice. *Journal of Consumer Research*, 21, 83-99.

Fischhoff, B., Slovic, P., Lichtenstein, S. 1978. Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 330 -344.

Harries, C., Harvey, N. 2000. Are Absolute Frequencies, Relative Frequencies, or Both Effective in Reducing Cognitive Biases? *Journal of Behavioral Decision Making*, 13, 431-444.

Henderson P.W., Peterson R.A. 1992. Mental accounting and categorization. *Organizational Behavior and Human Decision Processes*, 51, 92-117.

Hirt, E.R., Castellan, N.J. Jr. 1988. Probability and category redefinition in the fault tree paradigm. *Journal of Experimental Psychology; Human Perception and Performance*, 14, 112-131.

Idsen, L. C., Krantz, D. H., Osherson, D., Bonini, N. 2001. The Relation between Probability and Evidence Judgment: An Extension of Support Theory. *Journal of Risk and Uncertainty*, 22, 3, 227-249.

Kahneman, D., Tversky, A. 1973. On the psychology of prediction. *Psychological Review*, 80, 251-273.

Kahneman, D., Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

Koehler, D. J. 1994. Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 461-469.

Rottenstreich, Y., Tversky, A. 1997. Unpacking, Repacking, and Anchoring: Advances in Support Theory. *Psychological Review*, 104, 2, 406-415.

Russo, J.E., Kolzow, K. 1994. Where is the fault in fault trees? *Journal of Experimental Psychology: Human Perception and Performance*, 20, 1, 17-32.

Smith E.E., Shafir, E., Osherson, D. 1993. Similarity, plausibility, and judgments of probability. *Cognition*, 49, 67-96.

Teigen, K. H. 1983. Studies in subjective probability III: The unimportance of alternatives. *Scandinavian Journal of Psychology*, 24, 97-105.

Tversky, A., Koehler, D. J. 1994. Support Theory: A Nonextensional Representation of Subjective Probability. *Psychological Review*, 101, 4, 547-567.

Van der Pligt J., Eiser J.R., Spears, R. 1987. Comparative judgements and preferences: The influence of the number of response alternatives. *British Journal of Social psychology*, 26, 269-280.

Van der Pligt J., Van Schie E.C.M., Hoevenagel R. 1998. Understanding and valuing environmental issues: The effects of availability and anchoring on judgment. *Zeitschrift fur Experimentelle Psychologie*, 45, (4), 286-302.

Van Schie, E.C.M., Van der Pligt, J. 1990. Influence diagrams and fault trees: The role of salience and anchoring. In *Contemporary issues in decision making*, K. Borcherding, O.I. Larichev, D.M. Messick (eds.). Amsterdam: North Holland, pp. 177-188.

Wright, G., Whalley, P. 1983. The supra-additivity of subjective probability. In *Foundations of Utility and Risk Theory with Applications*, B. P. Stigum, F. Wenstop (eds.). Reidel Publishing Company, pp. 233-244.

EXHIBITS

	Competing Causes			
	Low Support		High Support	
	Diseases	Cancer	Diseases	Cancer
Competing causes				
Overall	.16*	.26*	.44*	.47*
Target causes				
Lung cancer	.43	.30	.27	.29
Infarctus	.36	.26	.29	.31
Road accident	.46	.36	.29	.33

Exhibit 1: Mean probability estimates of causes of death between the competing support conditions and the type of competing causes conditions. * mean sum of the probabilities assigned to the competing causes.

	Competing Causes	
	Low Support	High Support
Competing causes		
Overall	.16*	.83*
Target causes		
Lung cancer	.47	.25
Diabetes	.13	.08

Exhibit 2: Mean probability estimates of causes of death between the competing support conditions * mean sum of the probabilities assigned to the competing causes.