

# Typicality and reasoning fallacies

ELDAR B. SHAFIR

Princeton University, Princeton, New Jersey

EDWARD E. SMITH

University of Michigan, Ann Arbor, Michigan

and

DANIEL N. OSHERSON

Massachusetts Institute of Technology, Cambridge, Massachusetts

The work of Tversky and Kahneman on intuitive probability judgment leads to the following prediction: The judged probability that an instance belongs to a category is an increasing function of the typicality of the instance in the category. To test this prediction, subjects in Experiment 1 read a description of a person (e.g., "Linda is 31, bright, ... outspoken") followed by a category. Some subjects rated how typical the person was of the category, while others rated the probability that the person belonged to that category. For categories like *bank teller* and *feminist bank teller*: (1) subjects rated the person as more typical of the conjunctive category (a *conjunction effect*); (2) subjects rated it more probable that the person belonged to the conjunctive category (a *conjunction fallacy*); and (3) the magnitudes of the conjunction effect and fallacy were highly correlated. Experiment 2 documents an *inclusion fallacy*, wherein subjects judge, for example, "All bank tellers are conservative" to be more probable than "All feminist bank tellers are conservative." In Experiment 3, results parallel to those of Experiment 1 were obtained with respect to the inclusion fallacy.

## THE TYPICALITY PREDICTION

Well-known studies by Tversky and Kahneman (e.g., 1974, 1980, 1983) reveal that intuitive estimates of probability do not conform to the standard probability calculus. To explain their findings, Tversky and Kahneman impute to subjects heuristics that are often useful for estimating probability but that sometimes yield normatively incorrect judgments. One such heuristic estimates the probability that individual  $i$  has property  $P$  in terms of how representative  $i$  is of  $P$ . Most empirical investigation of the representativeness heuristic has involved instances and categories. In this case, the representativeness of an instance in a category reduces to the typicality of that instance in the category (roughly, how good an example the instance is of the category; see Smith & Medin, 1981, for a discussion). Tversky and Kahneman's hypothesis thus gives rise to the following prediction:

**TYPICALITY PREDICTION:** Subjects' estimates of the probability that instance  $i$  belongs to category  $C$  is an increasing function of the typicality of  $i$  in  $C$ .

The most striking evidence in support of this prediction arises in the context of the *conjunction fallacy*,

wherein people judge that an instance is more likely to be a member of a conjunctive category than of one of its constituent categories. For example, when presented with a description of a young woman, Linda, "who is 31 ... outspoken ... and concerned with issues of discrimination and social justice," subjects judge it more probable that Linda is a feminist bank teller than a bank teller (Tversky & Kahneman, 1980, 1983; see also Morier and Borgida, 1984, and Wells, 1985). According to Tversky and Kahneman, this fallacy arises because subjects assess Linda to be more typical of *feminist bank teller* than of *bank teller*, and then let this assessment influence their probability judgments; in support of this interpretation, Tversky and Kahneman (1983) report that a majority of subjects rated Linda as more typical of *feminist bank teller* than of *bank teller*. Tversky and Kahneman obtained parallel results for a second item involving the conjunctive category *accountant who plays jazz for a hobby*. Findings reported in Kahneman and Tversky (1973) and Pennington (1982) can also be interpreted as supporting the typicality prediction.

The foregoing studies are suggestive, but because each of them involved so few items, they provide only fragmentary evidence for the typicality prediction. To frame a more systematic test, we rely on the following definitions: Let a conjunctive category  $A\&B$  and an instance  $i$  be given.

The judged typicality of  $i$  in  $A\&B$  minus the judged typicality of  $i$  in  $B$  is termed the *conjunction effect for  $A\&B$  with respect to  $i$  and  $B$* .

Preparation of this manuscript was supported by US Public Health Services Grant MH 37208 to E. E. Smith and National Science Foundation Grant 870544 to D. Osherson. Correspondence may be addressed to Eldar Shafir, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544.

The judged probability that  $i$  belongs to  $A \& B$  minus the judged probability that  $i$  belongs to  $B$  is termed the *conjunction fallacy for  $A \& B$  with respect to  $i$  and  $B$* .

Then, to support the typicality prediction, we expect a positive correlation between

1. the conjunction effect for  $A \& B$  with respect to  $i$  and  $B$ , and
2. the conjunction fallacy for  $A \& B$  with respect to  $i$  and  $B$

across categories  $A, B$  and instances  $i$ . The purpose of our first experiment was to test this prediction.

To refine our test of the typicality prediction, the items included in the first experiment varied with respect to the compatibility of the categories  $A$  and  $B$  in the conjunction  $A \& B$ . For example, *feminist social worker* is a "compatible" conjunction, inasmuch as typical feminists share numerous properties with typical social workers. In contrast, *feminist bank teller* is an "incompatible" conjunction, inasmuch as typical feminists share few properties with typical bank tellers. Previous research (e.g., Smith, Osherson, Rips, & Keane, 1988) has shown that for an instance that is typical of  $A$ , the conjunction effect is larger for incompatible conjunctions  $A \& B$  than for compatible conjunctions (with respect to the instance and the category  $B$ ). To illustrate this finding, consider Linda as described above and the following four categories:

- a. feminist bank teller
- b. bank teller
- c. feminist social worker
- d. social worker

It can be shown that the difference in Linda's rated typicality for concepts (a) and (b) is greater than that for (c) and (d). In other words, the conjunction effect for (a) with respect to Linda and (b) is greater than the conjunction effect for (c) with respect to Linda and (d). According to the typicality prediction, the conjunction fallacy should show the same pattern.

In our second experiment, we demonstrate a new reasoning fallacy, called the *inclusion fallacy*, and we then show in Experiment 3 that it too is related to typicality. In the final section, we speculate about the nature of typicality judgments and consider why they are an imperfect guide to objective probability.

### Experiment 1 Test of the Typicality Prediction

One group of subjects made typicality judgments about personality descriptions, and another group made probability judgments about these same descriptions.

#### Method

**Materials.** Fourteen instances (i.e., personality descriptions) were generated, each associated with four categories (two of them conjunctive, two of them simple). Here is an illustrative instance with its associated categories:

Instance:

Jack began his job immediately after completing high school. He frequently talks on his CB radio, and goes to sporting events when he can.

Categories:

1. truck driver who plays softball for a hobby;
2. plays softball for a hobby;
3. truck driver who watches birds for a hobby;
4. watches birds for a hobby.

Notice that Jack is typical of *truck driver*, and that (1) is a compatible conjunction in the sense defined earlier, whereas (3) is an incompatible conjunction.

These instances and categories give rise to 28 triples, each triple consisting of an instance, a conjunctive category  $A \& B$ , and a constituent category  $B$ . Table 1 presents all 28 of these triples. In the table, each instance forms part of two triples: one triple includes an incompatible conjunction, the other a compatible conjunction.

We now describe the items used in Experiment 1, starting with those presented to subjects judging typicality. Each triple in Table 1 gives rise to two items—namely, the instance paired with category  $A \& B$ , and the instance paired with category  $B$ . The 28 triples thus yield a total of 56 items, such as:

Item:

Jack began his job immediately after completing high school. He frequently talks on his CB radio, and goes to sporting events when he can.

People who play softball for a hobby.

Each subject saw 28 of these items, broken down as follows: 7 instances paired with compatible categories like (1) and (2) above, and the other 7 instances paired with incompatible categories like (3) and (4) above. Counterbalancing over subjects assured that each instance was rated equally often with respect to compatible versus incompatible categories.

The items presented to subjects rating probability were identical to the typicality items except for the following change of wording. In place of categories such as "People who play softball for a hobby," we substituted statements, such as "Jack plays softball for a hobby."

**Procedure.** For both typicality and probability tasks, booklets were assembled, each page of which contained a single item. Each instance occurred once in the first half of the booklet, paired with either a conjunction or its constituent, and once in the second half of the booklet, paired with the counterpart to the first. Within each half-booklet, equal numbers of conjunctive and constituent categories occurred, and the order of items was randomized. The order of the two halves was counterbalanced across subjects.

Before receiving their booklets, subjects read an instruction page. The subjects who judged typicality were told that following each description of a particular person they would find the name of some category of people. Their task was to rate how typical the person described was of that category. Ratings were made on a scale between 0 and 1, where 1 indicated maximal typicality. The subjects who judged probability were told that following the description of a particular person, they would find a statement about that person. Their task was to rate the probability that the statement was true. Probabilities were rated between 0 and 1. The subjects in both tasks were asked to work through their booklets at their own rates and without referring back to earlier problems. Virtually all subjects took less than half an hour to complete the booklet.

Thus, for each triple in Table 1, our procedure provided ratings for: (1) the typicality of the instance in the category  $A \& B$ ; (2) the typicality of the instance in the category  $B$ ; (3) the probability that the instance belongs to the category  $A \& B$ ; and (4) the probability that the instance belongs to the category  $B$ .

**Subjects.** The subjects were 110 MIT and University of Michigan undergraduate volunteers, recruited from a variety of classes

**Table 1**  
**Conjunction Effects and Fallacies (Averaged over Subjects)**

| Instance   | Type | Conjunctive Category  | Constituent Category  | Effect   |          | Fallacy  |          |
|--|------|---|---|----------|----------|----------|----------|
|  |      |   |   | <i>i</i> | <i>c</i> | <i>i</i> | <i>c</i> |
| 1) Linda was a philosophy major. She is bright and concerned with issues of discrimination and social justice.   | i    | feminist bankteller<br>(typ=.439; prob=.401)  | bankteller<br>(typ=.264; prob=.241)                             | .18      |          | .16      |          |
|  | c    | feminist teacher<br>(typ=.733; prob=.601)   | teacher<br>(typ=.640; prob=.533)                                |          | .09      |          | .07      |
| 2) Jack began his job immediately after completing high school. He frequently talks on his CB radio, and goes to sporting events when he can.                      | i    | truck driver who watches birds for a hobby<br>(typ=.340; prob=.274)                   | watches birds for a hobby<br>(typ=.125; prob=.173)              | .22      |          | .10      |          |
|  | c    | truck driver who plays softball for a hobby<br>(typ=.773; prob=.707)                  | plays softball for a hobby<br>(typ=.664; prob=.623)             |          | .11      |          | .08      |
| 3) Hazel is quiet and somewhat shy, but she enjoys working with children. She always enjoyed school, particularly classes that had to do with crafts.              | i    | nurse who races bicycles for a hobby<br>(typ=.252; prob=.226)                         | races bicycles for a hobby<br>(typ=.177; prob=.160)             | .08      |          | .07      |          |
|  | c    | nurse who likes swimming for a hobby<br>(typ=.445; prob=.344)                         | likes swimming for a hobby<br>(typ=.294; prob=.355)             |          | .15      |          | -.01     |
| 4) Ted is 42 years old and loves chemistry. He reads a number of scientific magazines and is very interested in current scientific research.                       | i    | professor who plays the drums for a hobby<br>(typ=.337; prob=.367)                    | plays the drums for a hobby<br>(typ=.133; prob=.266)            | .20      |          | .10      |          |
|  | c    | professor who plays chess for a hobby<br>(typ=.716; prob=.686)                        | plays chess for a hobby<br>(typ=.623; prob=.598)                |          | .09      |          | .09      |
| 5) Jim is 40 years old and slightly overweight. He loves good wines and has an extensive cook-book library.  | i    | professional chef who boxes for a hobby<br>(typ=.289; prob=.269)                      | boxes for a hobby<br>(typ=.121; prob=.202)                      | .17      |          | .07      |          |
|  | c    | professional chef who listens to classical music for a hobby<br>(typ=.729; prob=.552) | listens to classical music for a hobby<br>(typ=.686; prob=.484) |          | .04      |          | .07      |
| 6) Richard is 50 years old. He loves his job, but is not very well liked by his colleagues. He is single, shy, and does not like to go out on social events.       | i    | engineer who plays volleyball for a hobby<br>(typ=.249; prob=.282)                    | plays volleyball for a hobby<br>(typ=.110; prob=.194)           | .14      |          | .09      |          |
|  | c    | engineer who collects stamps for a hobby<br>(typ=.604; prob=.550)                     | collects stamps for a hobby<br>(typ=.675; prob=.683)            |          | -.07     |          | -.13     |
| 7) Julie has a degree in physical education. Ever since childhood she has been in excellent physical shape. She loves the outdoors, and has an active social life. | i    | librarian who teaches aerobics in the evenings<br>(typ=.389; prob=.377)               | librarian<br>(typ=.161; prob=.152)                              | .23      |          | .22      |          |
|  | c    | ski instructor who teaches aerobics in the evenings<br>(typ=.772; prob=.578)          | ski instructor<br>(typ=.799; prob=.544)                         |          | -.03     |          | .03      |

Note—*i* = incompatible conjunction. *c* = compatible conjunction.

Table 1 (Continued)

| Instance  | Type     | Conjunctive Category   | Constituent Category                              | Effect   |          | Fallacy  |          |
|---|----------|--|---|----------|----------|----------|----------|
|   |          |  |   | <i>i</i> | <i>c</i> | <i>i</i> | <i>c</i> |
| 8) Laura was an English major. She loves going to the theater and reads a lot of poetry. She is also very interested in French literature.                | <i>i</i> | hair dresser who writes short stories for a hobby (typ=.383; prob=.252)            | hair dresser (typ=.208; prob=.188)                | .18      |          | .06      |          |
|   | <i>c</i> | works for a law firm and writes short stories for a hobby (typ=.564; prob=.516)    | works for a law firm (typ=.400; prob=.407)        |          | .16      |          | .11      |
| 9) John is 45 years old and has a degree in political science. He was a colonel in the army, and is fluent in Russian and Arabic.                         | <i>i</i> | works for the state department and grows flowers for a hobby (typ=.527; prob=.471) | grows flowers for a hobby (typ=.195; prob=.310)   | .33      |          | .16      |          |
|   | <i>c</i> | works for the state department and plays tennis for a hobby (typ=.694; prob=.559)  | plays tennis for a hobby (typ=.538; prob=.412)    |          | .16      |          | .15      |
| 10) Randy is 27 years old. He just graduated from art school, and loves going to museums and exhibitions. He dresses flamboyantly on social occasions.    | <i>i</i> | bus driver who paints in his free time (typ=.375; prob=.314)                       | bus driver (typ=.144; prob=.172)                  | .23      |          | .14      |          |
|   | <i>c</i> | photographer who paints in his free time (typ=.671; prob=.684)                     | photographer (typ=.668; prob=.642)                |          | 0        |          | .04      |
| 11) Carol is 34 years old and very ambitious. She is fluent in French, German and Spanish, and is interested in current political events.                 | <i>i</i> | foreign correspondent who knits for a hobby (typ=.559; prob=.580)                  | knits for a hobby (typ=.282; prob=.315)           | .28      |          | .27      |          |
|   | <i>c</i> | foreign correspondent who reads poetry for a hobby (typ=.750; prob=.519)           | reads poetry for a hobby (typ=.540; prob=.452)    |          | .21      |          | .07      |
| 12) Ralph graduated at the top of his class. He was the president of the student council, as well as the editor of the school's paper.                    | <i>i</i> | construction worker and president of a local labor-union (typ=.340; prob=.249)     | construction worker (typ=.078; prob=.131)         | .26      |          | .12      |          |
|   | <i>c</i> | lawyer and president of a local labor-union (typ=.575; prob=.509)                  | lawyer (typ=.779; prob=.707)                      |          | -.20     |          | -.20     |
| 13) Mike is 19 years old. He dropped out of high school and has been working since. He loves riding his motorcycle and going to football games.           | <i>i</i> | car mechanic who plays the flute for a hobby (typ=.291; prob=.339)                 | plays the flute for a hobby (typ=.116; prob=.180) | .18      |          | .16      |          |
|   | <i>c</i> | car mechanic who bowls for a hobby (typ=.668; prob=.503)                           | bowls for a hobby (typ=.512; prob=.453)           |          | .16      |          | .05      |
| 14) Helen is 26 years old. She is a very social person, but is not interested in politics. She loves buying new clothes and looking at women's magazines. | <i>i</i> | graduate student who frequently goes to fashion shows (typ=.423; prob=.439)        | graduate student (typ=.229; prob=.392)            | .19      |          | .05      |          |
|   | <i>c</i> | secretary who frequently goes to fashion shows (typ=.679; prob=.633)               | secretary (typ=.670; prob=.629)                   |          | .01      |          | 0        |

Average Effect or Fallacy: .21 .06 .13 .03

Number of Descriptions Showing Effect or Fallacy: 14 10 14 10

and paid for their participation. Fifty-four participated in the typicality task, and 56 in the probability task.

## Results

Conjunction effects were computed by subtracting the typicality rating of the instance with respect to the constituent *B* from that of the instance with respect to the conjunction *A&B*. A positive difference indicates a conjunction effect. Likewise, conjunction fallacies were computed by subtracting the probability rating of the instance with respect to the constituent *B* from that of the instance with respect to the conjunction *A&B*. A positive difference indicates a conjunction fallacy.

Any given subject made ratings relevant to only 14 of the 28 triples. Every subject produced some of the 14 possible conjunction effects or fallacies. The subjects who rated typicality produced an average of 7.9 effects, and those who rated probability produced an average of 7.5 fallacies. Table 1 presents the average effects and fallacies for each triple. Negative and zero numbers indicate the nonoccurrence of effects/fallacies (on the average).

It can be seen from Table 1 that every instance yielded conjunction effects and fallacies with respect to incompatible conjunctions, and that both effects and fallacies were reduced in frequency and magnitude with respect to compatible conjunctions. For triples with incompatible conjunctions, the average conjunction effect is .21 [which differs significantly from 0;  $t(13) = 12.5, p < .01$ ], while the average conjunction fallacy is .13 [which also differs significantly from 0;  $t(13) = 7.46, p < .01$ ]. For triples with compatible conjunctions, the average conjunction effect and fallacy are only .06 and .03, respectively (neither of which differs significantly from zero,  $p > .05$  in both cases). The average conjunction effect was significantly greater for triples with incompatible conjunctions than for triples with compatible conjunctions [ $t(13) = 2.89, p < .01$ ], as was the average conjunction fallacy [ $t(13) = 2.89, p < .01$ ]. The results reported in this paragraph remain significant when subjects rather than items serve as the units of analysis. As noted above, the foregoing pattern of results supports the typicality prediction.<sup>1</sup>

More direct support for the typicality prediction comes from the positive correlation between the magnitudes of the conjunction effects and fallacies. Thus, each of the 28 triples gives rise to both a conjunction effect and a conjunction fallacy. The magnitudes of these effects and fallacies correlate at .83 ( $p < .01, N = 28$ ). Consideration of just the triples with compatible conjunctions yields a correlation of .81 ( $p < .05, N = 14$ ). Consideration of just the triples with incompatible conjunctions yields a correlation of .58 ( $p < .05, N = 14$ ).

## THE SECOND TYPICALITY PREDICTION

Consider two categories *C'* and *C*. Suppose that a given subject judges every member of *C'* to be typical in *C*. Then, according to the typicality prediction, this subject ought to judge every member of *C'* to be a probable member of *C*. We are led in this way to a new prediction that

is a corollary of the original typicality prediction. It can be formulated as follows:

SECOND TYPICALITY PREDICTION: Subjects' estimates of the probability that every member of category *C'* is a member of category *C* is an increasing function of the typicality of members of *C'* in *C*.

Thus, according to the second typicality prediction, subjects' estimates of the probability that every bank teller is conservative depends on the typicality of bank tellers in the category *conservative*. In this example, the noun phrase "bank tellers" corresponds to *C'* of the second typicality prediction, whereas the adjective "conservative" corresponds to *C*.

Now suppose that subjects judge bank tellers, but not feminist bank tellers, to be typically conservative. Then they would be predicted to judge

Every single bank teller (in a particular bank) is conservative.  
as more probable than

Every single feminist bank teller (in that bank) is conservative.

Such a judgment violates the probability calculus inasmuch as the set of bank tellers includes its feminist subset. We call such an error an *inclusion fallacy*, and document its existence below. Notice that to commit the inclusion fallacy, subjects must assign a lower probability to a statement involving a conjunctive category *A&B* than to a parallel statement involving its constituent *A*. This pattern of probability attributions is the reverse of what occurs in the conjunction fallacy.

Our test of the second typicality prediction proceeds in two stages. First, we document the existence of the inclusion fallacy; this is the purpose of Experiment 2. Then, in Experiment 3, we replicate the fallacy and show that its strength can be predicted from typicality judgments, in conformity with the second typicality prediction.

## Experiment 2

### Prevalence of the Inclusion Fallacy

#### Method

**Materials.** Five adjectives were generated, each associated with one category and three subcategories. For example, one adjective is *conservative*; it is associated with the category *lawyers* and the subcategories *young lawyers*, *labor-union lawyers*, and *inexperienced lawyers*. We chose the adjectives to be more typical of the category than of some of the subcategories.

These adjectives and categories give rise to 15 triples, each triple consisting of a category *B*, a subcategory *A&B*, and an adjective-category *C*. Table 2 presents all 15 of these triples.

The inclusion items that were judged by subjects each had the frame: Every single \_\_\_\_\_ in the group is \_\_\_\_\_. The first slot was filled by either a category (e.g., *lawyer*) or a subcategory (e.g., *young lawyer*) drawn from a given triple, whereas the second slot was filled by the adjective (*conservative*) from that same triple. In this way, the 15 triples give rise to a total of 20 items (namely, 5 involving categories and 15 involving subcategories). Subjects rated the probability of all 20 items, which were presented on a single page in random order (except that occurrence of items containing the same adjective were separated by four other items).

**Procedure.** The rating task was preceded by the following tutorial, which was intended to ensure that the subjects understood the logical meaning of "every single."

**Table 2**  
**Inclusion Effects and Fallacies (Averaged over Subjects), Separately for Each Category, in Experiment 3**

| Subcategory ( <i>A&amp;B</i> )                | Category ( <i>B</i> )                 | Adjective Category ( <i>C</i> ) | Fallacy | Effect | Fallacy (Experiment 2) |
|---|---------------------------------------|---------------------------------|---------|--------|------------------------|
| labor-union<br>(typ=2.12;<br>prob=.166)       | lawyer<br>(typ=3.04;<br>prob=.204)    | conservative                    | .038    | .92    | .131                   |
| inexperienced<br>(typ=2.44;<br>prob=.217)     |                                       |                                 | -.014   | .60    | .039                   |
| young<br>(typ=2.42;<br>prob=.234)             |                                       |                                 | -.030   | .62    | .054                   |
| intelligent<br>(typ=2.82;<br>prob=.368)       | actor<br>(typ=3.44;<br>prob=.369)     | handsome                        | .001    | .62    | -.013                  |
| little-known<br>(typ=2.22;<br>prob=.296)      |                                       |                                 | .073    | 1.22   | .043                   |
| old<br>(typ=2.10;<br>prob=.349)               |                                       |                                 | .020    | 1.34   | .041                   |
| unsuccessful<br>(typ=1.98;<br>prob=.295)      | athlete<br>(typ=3.74;<br>prob=.500)   | strong                          | .205    | 1.76   | .292                   |
| woman<br>(typ=2.88;<br>prob=.501)             |                                       |                                 | -.001   | .86    | .027                   |
| thin<br>(typ=2.42;<br>prob=.372)              |                                       |                                 | .128    | 1.32   | .178                   |
| religious-studies<br>(typ=2.24;<br>prob=.259) | professor<br>(typ=2.60;<br>prob=.228) | liberal                         | -.031   | .36    | -.026                  |
| senior<br>(typ=2.10;<br>prob=.235)            |                                       |                                 | -.008   | .50    | .023                   |
| wealthy<br>(typ=1.96;<br>prob=.225)           |                                       |                                 | .003    | .64    | .036                   |
| liberal<br>(typ=2.56;<br>prob=.381)           | banker<br>(typ=3.44;<br>prob=.409)    | wealthy                         | .028    | .88    | .122                   |
| young<br>(typ=2.22;<br>prob=.294)             |                                       |                                 | .115    | 1.22   | .145                   |
| small-town<br>(typ=2.10;<br>prob=.364)        |                                       |                                 | .046    | 1.34   | .065                   |

Note—The corresponding inclusion fallacies of Experiment 2 are provided in the rightmost column for purposes of comparison. typ = average typicality judgment (on a scale from 1 to 4). prob = average probability judgment.

In this questionnaire we are interested in people's estimates of probabilities. To clarify the notion of probability, we begin with some examples intended to distinguish between probabilities and proportions.

Example 1:

(a) What is the probability that every single car on the Massachusetts Turnpike during the 5 o'clock rush hour is American-made? (Please note that "every single car is American-made" means that there is no car on that road that is not American-made.)

(b) What proportion of the cars on the Massachusetts Turnpike during the 5 o'clock rush hour are American-made?

Notice that while the answer to question (b) may be quite high, the answer to question (a) is low.

Example 2:

(a) What is the probability that every single musician in the Chicago Symphony Orchestra is male?

(b) What proportion of the musicians in the Chicago Symphony Orchestra is male?

Again, notice that the two questions are very different, and that the answers to them can also be quite different.

The subjects' average responses to the first example were .16 for the probability question versus .60 for the proportion question. Average responses to the second example were .22 for the probability question versus .66 for the proportion question. These estimates indicate that our subjects understood the meaning of "every single."

Directly following the tutorial, the subjects were given the following scenario:

Suppose that you are conducting a health survey and want to interview the largest variety of people possible. You walk into a shopping mall, hoping to find people of various ages, professions, social and economic classes, political views, etc. You hope, for example, that the accountants you meet will not all be the same age, the carpenters will not all have the same political views, etc.

In fact, in the group that you interview you encounter, among others: young bankers, liberal bankers, and small-town bankers; young lawyers, inexperienced lawyers, and labor-union lawyers; senior professors, wealthy professors, and religious-studies professors; unsuccessful athletes, thin athletes, and women athletes; and old actors, intelligent actors, as well as little-known actors.

The types of people listed in the above scenario were all the subcategories that appeared later in the probability and typicality items. Notice that the scenario sufficiently limits the groups involved so that a statement of the form "Every single \_\_\_\_\_ in the group is \_\_\_\_\_" has some nonnegligible probability of being true.

The subjects were then asked to rate the probability of the outcomes described in the 20 probability items (e.g., "Every single lawyer in the group is conservative"). The subjects worked through these items at their own speeds (approximately 15 min), and they were free to refer back to earlier items (which were on the same page). Immediately following the probability ratings, the subjects were asked to write the shopping-mall scenario from memory, in verbatim form as much as possible.

**Subjects.** The subjects were 40 University of Michigan undergraduate volunteers, recruited from a variety of classes and paid for their participation.

## Results

The following definition will facilitate the statement of results: Let a conjunctive category  $A\&B$  and another category  $C$  be given (e.g.,  $A = \text{young}$ ,  $B = \text{lawyer}$ ,  $C = \text{conservative}$ ). Then the judged probability that every member of  $B$  belongs to  $C$ , minus the judged probability that every member of  $A\&B$  belongs to  $C$ , is termed the *inclusion fallacy for  $A\&B$  with respect to  $C$  and  $B$* .

For each of the 15 triples figuring in the experiment, inclusion fallacies were computed by subtracting the prob-

ability rating associated with the subcategory and adjective (e.g., the rating for "Every single young lawyer in the group is conservative") from the rating associated with the category and adjective (e.g., "Every single lawyer in the group is conservative"). The sixth column of Table 2 presents the average fallacy associated with each triple. In the table, a positive difference indicates an inclusion fallacy. All subjects committed the inclusion fallacy, with an average of 10.1 fallacies per subject (out of a possible 15). Every triple yielded a fallacy for some subjects, with 13 out of the 15 triples yielding a fallacy when averaged over all subjects. The average fallacy over all 15 triples was .08 (on a scale from 0 to 1), which differs significantly from 0 [ $t(14) = 3.6$ ,  $p < .03$ ].

The prevalence of inclusion fallacies in the data reported above is unlikely to result from faulty memory for the shopping-mall scenario. Postexperimental interviews and inspection of subjects' attempts at verbatim recall provide informal evidence against such an interpretation of our findings. Evidence of a more objective nature was obtained as follows: Two readers each independently ranked the 40 recalled scenarios (1 for each subject) in terms of their fidelity to the original instructions. The Pearson rank-order correlation between the two rank orderings is .83. The faulty-memory hypothesis suggests an inverse relationship between the quality of a subject's recall and the number of fallacies committed. Thus, we split each set of rankings into two groups: better than median recall, and worse than median recall.  $T$  tests were performed for differences in the overall number of fallacies committed by the two groups. The results were the same for both rank orders:  $t(38) = 0.75$ , which is not statistically significant ( $p < .46$ ). We also determined the rank-order correlation between goodness of recall and number of fallacies committed. The correlation for the two rank orders are .01 and  $-.10$ , respectively.

## Experiment 3

### Test of the Second Typicality Prediction

To frame a systematic test of the second typicality prediction, we rely on the following definitions: Let a conjunctive category  $A\&B$  and another category  $C$  be given (e.g.,  $A = \text{young}$ ,  $B = \text{lawyer}$ ,  $C = \text{conservative}$ ).

The judged typicality of an arbitrary member of  $B$  in  $C$ , minus the judged typicality of an arbitrary member of  $A\&B$  in  $C$ , is termed the *inclusion effect for  $A\&B$  with respect to  $C$  and  $B$* .

The judged probability that every member of  $B$  belongs to  $C$ , minus the judged probability that every member of  $A\&B$  belongs to  $C$ , is termed the *inclusion fallacy for  $A\&B$  with respect to  $C$  and  $B$* .

Then, to support the second typicality prediction, we expect a positive correlation between

1. the inclusion effect for  $A\&B$  with respect to  $C$  and  $B$ , and
2. the inclusion fallacy for  $A\&B$  with respect to  $C$  and  $B$

across categories *A*, *B*, and *C*. The purpose of Experiment 3 was to test this prediction. Unlike the subjects in Experiment 1, the subjects in the present study made both typicality and probability judgments.

### Method

**Materials.** The five adjectives employed in Experiment 2 were again used in Experiment 3. Each adjective was associated with the same category and subcategories as before, yielding the same 15 triples shown in Table 2. The 15 triples give rise to 20 probability items, as described in Experiment 2. These same 20 items were employed for the probability task of Experiment 3. As before, subjects rated the probability of all 20 items, which were presented on a single page in random order (except that occurrences of items using the same adjective were separated by four other items).

We now consider items used in the typicality task. Each of the five adjectives figuring in the triples, along with its category and subcategories, were presented to subjects for typicality rating in the form exemplified as follows.

*Conservative:*

lawyer—  
young lawyer—  
labor-union lawyer—  
inexperienced lawyer—

The order in which the adjectives appeared was counterbalanced over subjects, as was the order of categories and subcategories used with a particular adjective.

**Procedure.** All subjects received both tasks. The probability task was performed first, preceded by the same tutorial as was used in Experiment 2, intended again to ensure that the subjects understood the logical meaning of "every single." The subjects' average responses to the first example were .0003 for the probability question versus .57 for the proportion question. Average responses to the second example was .02 for the probability question versus .72 for the proportion question.

Directly following the tutorial, the subjects were given the same shopping-mall scenario as in Experiment 2, and then asked to rate the probability of the outcomes described in the 20 probability items. Following a 10-min break, the subjects received the typicality task. They were asked to rate on a scale from 1 to 4 how typical each of the five adjectives was of the categories and subcategories listed below it (cf. the form illustrated above). A "1" indicated that the property was not typical, and a "4" indicated that it was extremely typical. Again the subjects worked through the page at their own speeds (approximately 10 min).

Thus, for each triple in Table 2, our procedure provided ratings for: (1) the typicality of an arbitrary member of category *B* in the category *C*; (2) the typicality of an arbitrary member of category *A&B* in the category *C*; (3) the probability that every member of category *B* belongs to category *C*; and (4) the probability that every member of category *A&B* belongs to category *C*.

**Subjects.** The subjects were 47 MIT undergraduate volunteers, recruited from a variety of classes and paid for their participation.

### Results

Regarding the probability task, the present findings replicate those of Experiment 2. For each triple, inclusion fallacies were computed as before, by subtracting the probability rating associated with the subcategory and adjective from the rating associated with the category and adjective. The fourth column of Table 2 presents the average fallacy associated with each triple. In the table, a positive difference indicates an inclusion fallacy. All but 2 subjects committed the inclusion fallacy, with an aver-

age of 5.6 fallacies per subject (out of a possible 15). Every triple yielded a fallacy for some subjects, with 10 out of the 15 pairs yielding a fallacy when averaged over all subjects. The average fallacy over all 15 triples was .04 (on a scale from 0 to 1), which differs significantly from 0 [ $t(14) = 2.2$ ,  $p < .05$ ].<sup>2</sup>

In the same way, for each triple, inclusion effects were computed by subtracting the typicality rating associated with the subcategory and adjective from the rating associated with the category and adjective. The fifth column of Table 2 presents the average effect associated with each triple; a positive difference indicates an inclusion effect. A positive difference was obtained in all 15 cases, its average magnitude being .94 on the 4-point scale [which differs significantly from 0;  $t = 9.2$ ,  $p < .01$ ].

To test the second typicality prediction, we correlated the magnitude of the inclusion effects and fallacies over all 15 triples. The obtained correlation was .87 ( $p < .01$ ,  $N = 15$ ).

Finally, note that for 12 out of the 15 triples, the average inclusion fallacy was greater in Experiment 2 than in Experiment 3. (See Table 2, columns 4 and 6.) This difference may be attributed to the different populations of subjects in the two studies: University of Michigan undergraduates in Experiment 2, MIT undergraduates in Experiment 3. The latter subjects are likely to be more mathematically prepared.

## SUMMARY AND CONCLUSIONS

### Summary

Our results support both typicality predictions. In both Experiments 1 and 3, probability judgments closely tracked typicality judgments. In Experiment 1, high correlations were obtained between the magnitudes of conjunction fallacies and effects for both the compatible and incompatible pairs, as well as over all pairs. In Experiments 2 and 3, a substantial number of inclusion fallacies were committed, and there was a high correlation in Experiment 3 between the magnitudes of the inclusion fallacies and inclusion effects.

We have obtained similar results using items drawn from the domains of fruits and vegetables rather than social categories. Thus, an instance described as "red, round, and mushy" is judged more likely to be a mushy apple than an apple. This is a conjunction fallacy. Moreover, the magnitude of this fallacy is positively correlated with the extent to which a "red, round, and mushy" object is judged to be more typical of *mushy apple* than of *apple*. Similarly, people judge Statement 1 to be more probable than Statement 2 (with respect to a particular supermarket containing both sour and nonsour apples):

1. Every single apple is red.
2. Every single sour apple is red.

This is an inclusion fallacy. As before, the magnitude of this fallacy is positively correlated with the extent to which *red* is judged to be more typical of *apple* than of *sour*



*apple*. In the domain of fruits and vegetables, however, the correlations between fallacies and effects—though still significant—are somewhat lower than those reported here for social categories.

**The Nature of Typicality Judgment**

Since typicality is a potent variable in predicting probability judgment, it is worth inquiring about the mechanisms underlying typicality judgment. In this subsection we consider three hypotheses about the manner in which subjects compute the typicality of instance *i* in category *C*.

**Hypothesis 1.** Subjects take typicality to be probability and hence take the typicality of instance *i* in category *C* to be the probability that *i* is a member of *C*. According to this hypothesis, our first and second typicality predictions state matters the wrong way round, since subjects rely on probability to compute typicality rather than the reverse. Hypothesis 1 thus provides a straightforward explanation of the correlation between typicality and probability found in Experiments 1-3. This hypothesis is discrepant, however, with two kinds of data. First, Tversky and Kahneman (1983) report cases in which judgments of typicality and probability diverge. For example, being divorced more than four times is considered more representative of Hollywood actresses than is voting democratic, yet the latter is judged more frequent among Hollywood actresses than the former. Since representativeness is related to typicality, and frequency to probability, we may interpret this finding as showing a divergence between typicality and probability judgment. If Hypothesis 1 were true, such cases could not exist. Second, it is a common observation that subjects who commit the conjunction fallacy— $\Pr(A\&B) > \Pr(B)$ —sometimes regret their judgment when reminded that *A* & *B* implies *B*. In contrast, we have never met a subject who regretted his or her judgment that Linda is more typical of *feminist bank teller* than of *bank teller* after being reminded that all feminist bank tellers are bank tellers. Again, such a difference is impossible on Hypothesis 1.

**Hypothesis 2.** Subjects mentally represent instances and categories as sets of features, and typicality judgment rests on comparison of these feature sets. There are several variants of this view (see Smith, 1989), but the simplest model is perhaps the following version of Tversky's (1977) contrast model: To assess, for example, the typicality of Linda in the category *bank teller*, the subject computes two weighted sums, namely: (1) the weighted sum of the features common to Linda's description and the category *bank teller*, and (2) the weighted sum of the features found in one of the latter two feature sets but not the other. These two weighted sums are then combined by a linear rule. Although typicality cannot be reduced to featural computations in every conceptual context (see Barsalou, 1985), such a mechanism is plausible for the categories figuring in the experiments reported here. Moreover, the reader may verify that, using any reasonable feature-weighting function, the present hypothesis correctly predicts the observed variation in typicality judgment reported in Experiments 1-3.

Featural computations like those embodied in Tversky's contrast model allow typicality to be partially quantified and then related to probability judgment. As discussed in Osherson (1990), this approach to probability judgment has considerable predictive power. Nonetheless, current featural models seem not to provide a complete account of probability judgment, since they cannot encode information about real-world constraints without additional principles governing feature weights. To see the problem, consider Margaret Thatcher and Jean Kirkpatrick. According to informal opinion, Thatcher is about as likely as Kirkpatrick to require bifocals sometime over the next 5 years, yet the latter is infinitely more likely than the former to be nominated for president of the United States. To predict these differences in probability using featural overlap, it is evident that features corresponding to nationality must be slightly weighted in the ocular context and greatly weighted in the political context. As yet, no general principles have been formulated for assigning weights to features in different contexts. Further examples and discussion are provided in Osherson, Smith, and Shafir (1986, Section 2.7). Since featural models represent at best an incomplete approach to probability, it may be useful to consider a third hypothesis about the relationship between typicality and probability judgment.

**Hypothesis 3.** Subjects take the typicality of instance *i* in category *C* to be the judged probability of something's being *i*-like, given that it belongs to *C*. Like the first hypothesis, Hypothesis 3 invokes probability judgment as underlying typicality. The two hypotheses differ in a crucial respect, however. Unlike in Hypothesis 1, the probabilities invoked in Hypothesis 3 are not the ones requested of subjects in the probability tasks of Experiments 1-3. To make this point clear, consider again the typicality of Hollywood actress in the categories *divorced more than four times* and *votes democratic*. The present hypothesis claims that these typicalities are given by:

- a.  $\Pr(X \text{ is a Hollywood actress} | X \text{ has divorced more than four times})$
- b.  $\Pr(X \text{ is a Hollywood actress} | X \text{ votes democratic}),$

where *X* is an arbitrary person. It is clear that (a) is likely to be judged as greater than (b), so the typicality facts are correctly predicted by Hypothesis 3. Hypothesis 1, in contrast, represents the two typicalities by:

- a'.  $\Pr(X \text{ has divorced more than four times} | X \text{ is a Hollywood actress})$
- b'.  $\Pr(X \text{ votes democratic} | X \text{ is a Hollywood actress}).$

Tversky and Kahneman's data show that (b') is rated as higher than (a'), hence Hypothesis 1 makes wrong predictions about typicality.

Hypothesis 3 combined with the first typicality prediction implies that subjects will commit the conjunction fallacy when evaluating:

- 1. a.  $\Pr(\text{Linda is a bank teller})$
- b.  $\Pr(\text{Linda is a feminist bank teller}),$

where Linda conforms to the usual description (see under

Typicality Prediction above). The deduction requires three steps. First, the typicality prediction implies that (1b) > (1a) if and only if (2b) > (2a):

2. a. the typicality of the instance Linda in the category *bank teller*
- b. the typicality of the instance Linda in the category *feminist bank teller*.

Second, Hypothesis 3 implies that (2b) > (2a) if and only if (3b) > (3a):

3. a. Pr( $X$  has Linda's described properties |  $X$  is a bank teller)
- b. Pr( $X$  has Linda's described properties |  $X$  is a feminist bank teller).

Finally, it is clear that most subjects would take (3b) to exceed (3a), which yields (2b) > (2a), hence (1b) > (1a), both documented empirically. The other conjunction fallacies appearing in Experiment 1 may be deduced in the same manner.

Similarly, Hypothesis 3 combined with the second typicality prediction implies that subjects will commit the inclusion fallacy when evaluating:

4. a. Pr(All lawyers are conservative)
- b. Pr(All young lawyers are conservative).

The deduction again requires three steps. The second typicality prediction implies that (4a) > (4b) if and only if (5a) > (5b):

5. a. the typicality of a lawyer-instance in the category *conservative*
- b. the typicality of a young-lawyer-instance in the category *conservative*.

Next, Hypothesis 3 implies that (5a) > (5b) if and only if (6a) > (6b):

6. a. Pr( $X$  is a lawyer |  $X$  is conservative)
- b. Pr( $X$  is a young lawyer |  $X$  is conservative).

Finally, it seems clear that most subjects would take (6a) to exceed (6b), which yields (5a) > (5b); hence (4a) > (4b), both documented empirically.

Hypothesis 3 portrays the subject's poor probability judgment as the result of calculating the wrong probability—namely, the probability used to compute typicality. Observe, for example, that (3a,b) are not equivalent to (1a,b), nor (6a,b) to (4a,b). Of course, Hypothesis 3 does not exclude the use of feature overlap in computing typicality, for the probabilities underlying typicality judgments may well be calculated partially on a featural basis.

### Concluding Remarks

We conclude with two comments about the inclusion fallacy. First, the prevalence of the inclusion fallacy argues against interpretations of the conjunction fallacy which

assume that conjunctions are favored over constituents because the former contain more information than the latter. For, in the inclusion fallacy, it is the less informative category that is erroneously favored over the more informative (conjunctive) subcategory. Second, the prevalence of the inclusion fallacy supports Tversky and Kahneman's (1983) claim that intuitive probability estimates are not extensional. An extensional estimate of the probability that your new neighbor is unmarried might be based on the probability that he is divorced plus the probability that he is widowed plus the probability that he has never married. People rarely seem to decompose a target event like this, and failure to decompose a category into its subcategories sets the stage for the inclusion fallacy.

### REFERENCES

- BARSALOU, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 629-654.
- KAHNEMAN, D., & TVERSKY, A. (1973). The psychology of prediction. *Psychological Review*, **80**, 237-251.
- MORIER, D. M., & BORGIDA, E. (1984). The conjunction fallacy: A task-specific phenomenon? *Personality & Social Psychology Bulletin*, **10**, 243-252.
- OSHERSON, D. (1990). Judgment. In D. Osherson & E. E. Smith (Eds.), *Invitation to cognitive science* (Vol. 3, pp. 55-58). Cambridge, MA: MIT Press.
- OSHERSON, D., SMITH, E. E., & SHAFIR, E. (1986). Some origins of belief. *Cognition*, **24**, 197-224.
- PENNINGTON, N. (1982, November). *Comments on Tversky and Kahneman's 'Intensional reasoning and the conjunction fallacy'*. Paper presented at the Annual Meeting of the Judgment/Decision Making Conference, Minneapolis, MN.
- SMITH, E. E. (1989). Concepts and induction. In M. Posner (Ed.), *Foundations of cognitive science* (pp. 502-526). Cambridge, MA: MIT Press.
- SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- SMITH, E. E., OSHERSON, D. N., RIPS, L. J., & KEANE, M. (1988). Combining prototypes: A selective modification model. *Cognitive Science*, **12**, 485-527.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.
- TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124-1131.
- TVERSKY, A., & KAHNEMAN, D. (1980). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- TVERSKY, A., & KAHNEMAN, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293-315.
- WELLS, G. L. (1985). The conjunction error and the representativeness heuristic. *Social Cognition*, **3**, 266-279.

### NOTES

1. The finding that conjunction effects and fallacies were greater for incompatible than compatible conjunctions is not due to a ceiling effect for the latter. The average ratings of compatible conjunctions in both the typicality and probability tasks (.67 and .57, respectively) never approached the ceiling of the permitted range (which was bounded between 0 and 1).

2. Our definition of an inclusion fallacy is conservative. A difference of 0 probability between category and subcategory is itself a fallacy, because subjects knew that the subcategory was properly included in the category. Thus, it is a fallacy to assign probability .5 to both "Every single lawyer is conservative" and "Every single young lawyer is conservative" when one knows of lawyers other than young ones. According

to this more liberal definition, all of our subjects committed the inclusion fallacy, yielding an average of 9.4 (out of a possible 15) fallacies per subject. Similar remarks apply to the results of Experiment 2.

(Manuscript received April 27, 1988;  
revision accepted for publication September 14, 1989.)