Probabilistic Dependence Between Events

Ruma Falk
Maya Bar-Hillel

Ruma Falk is teaching Probability and Statistics at The Hebrew University of Jerusalem in the Department of Psychology and in The School of Social Work. She received her Ph.D. from The Hebrew University for a thesis on "The Perception of Randomness." Her research interests are in probabilistic thinking in children and adults. Dr. Falk spent the academic year 1979/80 at Decision Research in Eugene, Oregon, where most of this paper was written.

Maya Bar-Hillel is a psychologist at The Hebrew University. She has written on philosophical, psychological, mathematical and educational aspects of probability and induction.

> *While investigating an armed robbery, officers Jones and Smith pursued independent lines of inquiry. Both obtained different testimony placing the suspect close to the scene of the crime at about the time it took place. Upon comparing notes, however, they had to drop charges, since the pooled testimony conclusively proved that the suspect could not possibly have been at the scene of the crime when it was committed.* Adapted from Tribe [9, p. 1367]

Two events, A and B, in a probability space (S, P) are customarily classified as being either independent [i.e., $P(A \cap B) = P(A)P(B)$, or $P(B \mid A) = P(B)$] or dependent [i.e., $P(A \cap B) \neq P(A)P(B)$, or $P(B \mid A) \neq P(B)$]. The relationship of dependence, however, lends itself very naturally to a further classification, namely $P(B \mid A) > P(B)$ versus $P(B \mid A) < P(B)$ [or $P(A \cap B) > P(A)P(B)$ versus $P(A \cap B) < P(A)P(B)$]. The former might be called "positive relevance" and the latter "negative relevance;" in this light, independence might also be called "irrelevance."

If the distinction between dependence and independence is deemed important, certainly the distinction between positive and negative relevance is even more so. Evidence tells us little if we know that it is relevant to some hypothesis, but do not know the direction in which it affects the probability. Nonetheless, the distinction between positive and negative relevance is so rarely encountered, that it even lacks an accepted label or term.

In the present paper we first suggest a threefold classification of dependence relationships between pairs of events, then point out some misconceptions concerning these relationships, and, lastly, speculate as to the reasons that it is not customarily employed.

Analyzing Dependence Relationships. We begin by introducing some notations and definitions. Consider only strictly uncertain events X, i.e., $0 < P(X) < 1$.

$$A \nearrow B \text{ if } P(B \mid A) > P(B); \quad A \searrow B \text{ if } P(B \mid A) < P(B); \quad A \perp B \text{ if } P(B \mid A) = P(B).$$

As shown in Table 1, this threefold classification describes the different ways in which knowledge of the occurrence of A can affect the probability of B. It can increase it ($\nearrow$); it can lower it ($\searrow$); it can leave it unchanged ($\perp$) (note that these relationships are not set-theoretical ones, but rather are induced by the probability function P). Surprisingly, perhaps, it turns out that these relationships do not behave in a systematic or "orderly" fashion, and fail to fulfill several natural expectations.

Table 1.
Threefold Classification of Probabilistic Dependence Relationships.

Formal relationship	$P(B \mid A) > P(B)$	$P(B \mid A) < P(B)$	$P(B \mid A) = P(B)$
Notation	$A \nearrow B$	$A \searrow B$	$A \perp B$
Name of relationship	A supports B	A weakens B	B is independent of A
Classification of dependence	Positive dependence	Negative dependence	Independence

Theorem 1. *The dependence relationships between events are symmetrical, i.e.,*
$A \nearrow B$ *if and only if* $B \nearrow A$
$A \searrow B$ *if and only if* $B \searrow A$
$A \perp B$ *if and only if* $B \perp A$.

The proof follows directly from the definitions, and we skip it here.

Theorem 1 allows us to use the directional name of the relationship (e.g., "A supports B") and the nondirectional classification ("A and B are positively dependent") interchangeably. This result should not be surprising, unless one tends to confuse conditionality with causality [5], [10]. In that case it may seem (erroneously) that $A \nearrow B$ when A is a cause of B, but that since B may not be a cause of A, the converse, $B \nearrow A$, need not hold (see Example 1 in the Appendix).

More intriguing is the following theorem:

Theorem 2. *The relationships* $\nearrow$, $\searrow$, *and* $\perp$ *are not transitive.*

A formal proof for independence and for negative dependence can be derived by a reductio ad absurdum argument from the symmetry and the nonreflexivity of these relationships. Since $A \searrow B$ implies $B \searrow A$ (and $A \perp B$ implies $B \perp A$), transitivity would result in the absurdity $A \searrow A$ (and $A \perp A$). The same argument cannot be applied to $\nearrow$, since A does support itself.

It is not surprising that $\searrow$ is not transitive. One might expect it, in fact, to be intransitive, since, in a manner of speaking, a negation of a negation is tantamount to an affirmation (see Example 2 in the Appendix). For analogous reasons, one might expect $\nearrow$ to be transitive. That it is not can, however, be demonstrated by a Venn diagram, as in Figure 1, which depicts three events such that A supports B, B supports C, but A does not support C. The graphic counterexample incorporated into Figure 1 might be even more illuminating if it were given a concrete interpretation as in the following example:

Let A, B, and C denote the following events, defined in the adult male population: A—having a full head of white hair; B—being over fifty years old; C—being completely bald. Then $A \nearrow B$, since there are proportionately more men over fifty among those with white hair than in the adult male population at large; $B \nearrow C$, since there are proportionately more bald men in the over-fifty age bracket than in the adult male population at large; but clearly $A \searrow C$, since men with full heads of white hair are not bald.

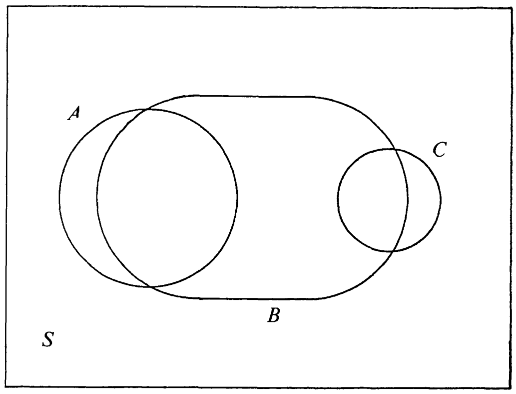

Figure 1.

A Venn diagram showing $A \nearrow B$, $B \nearrow C$ & $A \searrow C$.

It is possible to relate properties of $\nearrow$ to properties of $\searrow$.

Theorem 3. *$A \nearrow B$ if and only if $A \searrow \bar{B}$.*

From this theorem, we can derive all others relating $\nearrow$ to $\searrow$, such as $A \nearrow B$ if and only if $\bar{A} \searrow B$, $A \nearrow B$ if and only if $\bar{A} \nearrow \bar{B}$, etc. Since these theorems enable us to form analogous theorems for $\nearrow$ and for $\searrow$, we shall address ourselves below to $\nearrow$ only.

The following is a list detailing properties that *do not* hold with respect to $\nearrow$, although one might expect them to.

Theorem 4. $A \nearrow C, B \nearrow C \not\Rightarrow A \cap B \nearrow C.$

Theorem 5. $A \nearrow C, B \nearrow C \not\Rightarrow A \cup B \nearrow C.$

Theorem 6. $C \nearrow A, C \nearrow B \not\Rightarrow C \nearrow A \cap B.$

Theorem 7. $C \nearrow A, C \nearrow B \not\Rightarrow C \nearrow A \cup B.$

Note that the same theorems apply to negative dependence and to independence as well.[1]

[1]This list is, of course, redundant, since Theorem 6 follows from 4, and 7 follows from 5 by symmetry. Less obvious is that 5 follows from 4 (and 7 from 6) by successive applications of a corollary of Theorem 3 (namely, $A \nearrow B$ if and only if $\bar{A} \nearrow \bar{B}$), in conjunction with DeMorgan's laws (i.e., the theorems relating $\cap$ to $\cup$ via complementation, e.g., $\overline{A \cap B} = \bar{A} \cup \bar{B}$).

False Expectations Concerning Dependence Relationships. Theorems 4 through 7 have some startling implications for inductive inference. For example, Theorem 4 means that even though two items of evidence may individually support some hypothesis C, their conjunction might not; in fact, it might even weaken it! Such vagaries of evidential relevance, neglected by mathematicians, have been studied primarily by philosophers (e.g., [3], [6], [8]).

The fact that philosophers have seen fit to warn against possible acceptance of these properties suggests that there is an intuitive tendency to do so. The readers may check their own intuitions to verify their appeal. Some informal results, which we gathered during a classroom survey, indicate that college students tend to believe that the properties do hold.[2] One subject, in explaining why he believed that $A \nearrow C$ and $B \nearrow C$ implies $A \cap B \nearrow C$, captured nicely what we believe to be a prevalent intuition, and we quote: "If A supports C and B supports C, then their conjunction does so *all the more*!"

Despite endorsing these relationships in the abstract, when subjects were given specific, concrete examples of the premises of Theorems 4 through 7, they readily drew the correct inference from the data even when it ran contrary to the conclusion expected in the abstract. This makes us feel that the best way to treat these relationships in the classroom is through examples (see Examples 3 and 4 in the Appendix).

One such example, due to Carnap [3, p. 382], appears in Table 2. Table 2 concerns ten people who are simultaneously distributed according to three dichotomous variables: Sex (M-male, F-female); Age (Y-young, O-old); Marital status (S-single, W-wed). A person is now drawn at random. It is easy to verify, for example, that $S \nearrow F$ and $O \nearrow F$ while $S \cup O \searrow F$, in support of Theorem 5. The reader may enjoy the exercise of finding counterexamples to the other properties in Table 2 as well.

Table 2.
Distribution of Ten People by Age, Sex and Marital Status
(after Carnap [3]).

	S (Single)	W (Wed)
Y (Young)	M, M	M, F, F
O (Old)	F, F, F	M, M

A less technical example, closer to the ways one might encounter $\nearrow$ in real-life reasoning, appears at the opening of this paper [9]. We now supply the details that resolve the apparent puzzle posed by this example.

The armed robbery, known to have taken at least fifteen minutes, was committed between 3:00 a.m. and 3:30 a.m. Jones learned E_1—that the suspect was seen in a car close to the scene of the crime at 3:10 a.m. Since the evidence places the suspect

[2] A more detailed description of our little experiment can be obtained by writing to the authors.

near the scene of the crime at the right time ("opportunity"), he reassessed his initial probability for G—the event that the suspect was involved in the robbery, $P(G \mid E_1) > P(G)$, i.e., $E_1 \nearrow G$. Smith learned that the suspect was seen in a bar close to the scene of the crime at 3:20 a.m. Following Jones' reasoning, he concluded that $E_2 \nearrow G$. Upon comparing notes, they realized that the conjunction of their investigative efforts essentially supplies the suspect with an alibi. Thus the initial probability of G drops, i.e., $E_1 \cap E_2 \searrow G$.

The way the occurrence of one event affects the probability of another event is a very important relationship for probabilistic inference. The disappointing irregularity of this relationship may help to explain why it has been largely ignored by mathematicians, and why standard texts in probability theory fail to define or discuss it.[3] There is just very little that can be systematically said about it. Nonintuitive theorems such as listed above are among the obstacles confronting philosophers who were interested in building a "calculus of induction," confirmation or relevance [3], [6].

Our impressions suggest that the student who has not explicitly studied the properties of dependence relationships will supply his or her own, often erroneous, intuitions regarding their behavior. In particular, without being cautioned, one may automatically assume that positive dependence is transitive and that it satisfies the properties claimed by Theorems 4–7 not to hold. These implicit assumptions are not so deeply entrenched, however, as to be impossible to extinguish. We believe that a discussion, accompanied by examples, will convince most students of their fallaciousness. This effort is worthwhile in spite of the fact that counterexamples to the false intuitions may be ecologically rare in "real life" or in scientific work. Usually, converging evidence is cumulative, and conjunctions support a tested hypothesis beyond the impact of each individual piece of evidence. Studying the characteristics of dependence relationships seems important not so much for its applied value, but rather for its theoretical importance and for its significance for the methodology of "convergent operationalism" [2 p. 129].

Probabilistic Versus Deductive Inference. Probabilistic support has often been looked upon as a degraded kind of deductive proof. That is to say, enhancing the probability of a hypothesis is regarded as a step toward proving it. This notion might well be the source of people's erroneous expectations. They may merely be attributing to $\nearrow$ the properties of the logical relationship of implication, denoted $\rightarrow$. Note that substituting $\rightarrow$ for $\nearrow$ necessitates replacing $\not\Rightarrow$ by $\Rightarrow$ in Theorems 4–7, since all four properties are true for logical implication. However, evidential support is not just a weak form of logical proof, and probabilistic inference is not a degraded form of deductive inference. The two are different systems* with different rules [1]. This, perhaps, is the most important lesson from our discussion. Table 3 classifies some simple rules of probabilistic and deductive inference with respect to their truth or falsity. The table should be read as follows: By substituting $\nearrow$ for R, we obtain a statement concerning positive dependence. By substituting $\rightarrow$ for R, we obtain a statement concerning logical implication. The truth value of these state-

[3] Nevertheless, a problem concerning $\searrow$ can be found in Holm [7, p. 348].

* $\cap$ and $\cup$ are used for both systems because their use in set theory is fully analogous to conjunction and disjunction in logic. In fact, historically one of the earliest systems of notation used $\cap$ and $\cup$ for conjunction and disjunction, respectively. In recent times, $\wedge$ and $\vee$ have been more commonly used for conjunction and disjunction.

Table 3.
Cross Classification of Rules Governing Logical Implication and Probabilistic Support[a].

		Logical Implication (L) $A \rightarrow B$	
		True	False
Probabilistic Support (P) $A \nearrow B$	True	ARA $ARB \Rightarrow \bar{B}R\bar{A}$	$ARB \Rightarrow BRA$ $ARB \Rightarrow \bar{A}R\bar{B}$
	False	$ARB \,\&\, BRC \Rightarrow ARC$ $ARC \,\&\, BRC \Rightarrow (A \cap B)RC$ $ARC \,\&\, BRC \Rightarrow (A \cup B)RC$	$ARB \Rightarrow A$ precedes B temporally

The relationship R should be replaced, in turn, by both $\rightarrow$ and $\nearrow$ and the truth value of the resultant proposition verified.

[a] Probabilistic support ($\nearrow$) can be replaced by probabilistic weakening ($\searrow$) everywhere, except for ARA, since $\searrow$ is *non*reflexive.

ments is then determined by their row or column classification, respectively. For instance, in the first line of the upper right cell of the table, we read $ARB \Rightarrow BRA$. This means that for probabilistic support $A \nearrow B \Rightarrow B \nearrow A$ is *true* (by row), while for logical implication $A \rightarrow B \Rightarrow B \rightarrow A$ is *false* (by column).

Note that the cell "False-False" is devoid of mathematical characteristics. The "temporally precedes" relationship was included for didactic reasons because of its psychological significance. We wished to emphasize that neither probabilistic support nor logical implication entails a cause-and-effect relationship. A may imply or support B even when A is the effect and B is the cause (as when A means "rain" and B means "clouds"), or when both are caused by C.

In the cell "False(P)-True(L)" we find rules that are true only for logical inference, and in the "True(P)-False(L)" cell we find rules that are true only for probabilistic inference. We have dwelt in this paper on people's inclination to expect the rules true for logical implication to hold for probabilistic inference. We have not considered the other side of the same coin, i.e., the fact that sometimes the dual generalization takes place. In other words, sometimes rules that apply to probabilistic support are (falsely) considered true with respect to logical implication. Thus, in the cell "True(P)-False(L)" we find two rules that students of Introductory Logic often find unintuitive. In fact, since the logical relationship of implication and fallacies therefrom have been extensively studied, the misconceptions in this cell even have standard names. They are called, respectively, "the Fallacy of Affirming the Consequent" and "the Fallacy of Denying the Antecedent" [4, p. 295]. These are not fallacies with respect to $\nearrow$. Thus, the misconceptions typical for both cells in the diagonal true/false indicate that people tend to ascribe the same properties to logical implications and to probabilistic support, as if they have a shared pool of intuitions about these two methods of inference.

An explicit discussion of dependence relationships could both highlight the distinction between these two modes of thought and alert the student to the pitfalls of generalizing from one system to the other.

Acknowledgement: The authors wish to thank Baruch Fischhoff and Sarah Lichtenstein for their helpful comments.

Appendix

Example 1. An urn contains two white balls and two black balls. We shake the urn thoroughly and blindly draw out two balls, one after the other, *without* replacement.

Let W_1 denote the event "the first ball drawn is white," and let B_2 denote: "the second ball drawn is black."

Clearly $W_1 \nearrow B_2$, since $P(B_2) = \frac{1}{2}$ whereas $P(B_2 \mid W_1) = \frac{2}{3}$. However, when asked about $P(W_1 \mid B_2)$, students often answer $\frac{1}{2}$, claiming that subsequent events cannot affect the outcomes of previous ones [5]. This causal reasoning erroneously implies that $W_1 \nearrow B_2$ but $B_2 \perp W_1$. Actually, however, the information about the subsequent event affects the uncertainty with regard to the previous event. Since the self-same black ball cannot be drawn twice $P(W_1 \mid B_2)$ also equals $\frac{2}{3}$, i.e., $B_2 \nearrow W_1$.

Example 2. The somewhat vague expression "a negation of a negation is tantamount to an affirmation" can be given a concise explication, as, for instance, below: Suppose three witnesses, X, Y, and Z, are giving the following testimonies:

X: "Y is untrustworthy."
Y: "Z is untrustworthy."
Z: "The butler did it."

Let A, B, and C be the following events:

A: X is telling the truth, i.e., Y is untrustworthy.
B: Y is telling the truth, i.e., Z is untrustworthy.
C: Z is telling the truth, i.e., the butler did it.

Now, $A \searrow B$ since

$$P(B \mid A) = P(Z \text{ is untrustworthy} \mid Y \text{ is untrustworthy})$$
$$< P(Z \text{ is untrustworthy}) = P(B).$$

Also $B \searrow C$ since

$$P(C \mid B) = P(\text{the butler did it} \mid Z \text{ is untrustworthy})$$
$$< P(\text{the butler did it}) = P(C).$$

However, $A \not\searrow C$, since indirectly X's testimony (which is assumed to be true) supports Z's trustworthiness (by discrediting Y's testimony regarding Z's untrustworthiness), i.e.,

$$P(C \mid A) = P(\text{the butler did it} \mid Y \text{ is untrustworthy}) > P(\text{the butler did it}) = P(C).$$

Example 3. In what follows, $A \perp C$, $B \perp C$, but $A \cap B \not\perp C$ (see Theorem 4).

We independently toss two dice, one white and one black. Consider the following events:

A—an even outcome on the white die.
B—an even outcome on the black die.
C—an even *sum*.

These three events satisfy: $P(C) = 1/2$; $P(C \mid A) = P(C \mid B) = 1/2$; i.e., $A \perp C$ and $B \perp C$.

However, $P[C \mid (A \cap B)] = 1 > P(C)$, i.e., $A \cap B \nearrow C$.

Example 4. Consider the following anecdote (related by Gardner [6, p. 121]):

> A research project shows that 3/5 of a group of patients taking a certain pill are immune to colds for five years, compared with only 2/5 in the control group who were given a placebo. A second project shows that 3/5 of a group receiving the pill were immune to tooth cavities for five years, compared with 2/5 who got the placebo. The combined statistics could show that twice as many among those who got the placebo are free for five years from both colds and cavities compared with those who got the pill.

Let us denote:
C—immune to colds ($\bar{C}$—not immune to colds).
T—immune to tooth cavities ($\bar{T}$—not immune to tooth cavities).

Suppose an experiment involved 10 patients, of whom 5 received the pill and 5 received the placebo. The following list of outcomes conforms to all the experimental results described above:

Pill	Placebo
$C \cap T$	$C \cap T$
$C \cap \bar{T}$	$C \cap T$
$C \cap \bar{T}$	$\bar{C} \cap \bar{T}$
$\bar{C} \cap T$	$\bar{C} \cap \bar{T}$
$\bar{C} \cap T$	$\bar{C} \cap \bar{T}$

Note that Pill$\nearrow C$ and Pill$\nearrow T$, while Pill$\searrow C \cap T$ (see Theorem 6).

REFERENCES

1. E. W. Adams, The Logic of Conditionals, Reidel, Dordrecht, Holland, 1975.
2. D. T. Campbell and D. W. Fiske, Convergent and discriminant validation by the multitrait-multimethod matrix, in D. N. Jackson and S. Messick (Editors), Problems in Human Assessment, McGraw-Hill, New York, 1967, pp. 124–132.
3. R. Carnap, Logical Foundations of Probability, Chapter 6, University of Chicago Press, Chicago, 1950.
4. I. M. Copi, Introduction to Logic, 5th ed., Macmillan, New York, 1978.
5. R. Falk, Revision of probabilities and the time axis, Proc. Third Internat. Confer. Psych. Math. Ed., Warwick, England, 1979, pp. 64–66.
6. M. Gardner, On the fabric of inductive logic and some probability paradoxes, Sci. Amer., 234 (1976) 119–122.
7. S. Holm, A collection of problems in probability and statistics, in L. Råde (Editor), The Teaching of Probability and Statistics, Almqvist and Wiksell, Stockholm, 1970.
8. W. C. Salmon, Confirmation, Sci. Amer., 228 (1973) 75–83.
9. L. H. Tribe, Trial by mathematics: Precision and ritual in the legal process, Harvard Law Review, 84 (1971) 1329–1393.
10. A. Tversky and D. Kahneman, Causal schemata in judgment under uncertainty, in M. Fishbein (Editor), Progress in Social Psychology, Erlbaum, Hillsdale, NJ, 1979.