Work in progress
Send comments to tshogenji@ric.edu

# The Degree of Epistemic Justification and the Conjunction Fallacy

**Tomoji Shogenji**

(Last revision made on August 27, 2008)

This paper describes a formal measure of epistemic justification motivated by the dual goal of cognition, which is to increase true beliefs and reduce false beliefs. From this perspective the degree of epistemic justification should not be the conditional probability of the proposition given the evidence, as it is commonly thought. It should be determined instead by the combination of the conditional probability and the prior probability. This is also true of the degree of incremental confirmation, and I argue that any measure of epistemic justification is also a measure of incremental confirmation. However, the degree of epistemic justification must meet an additional condition, and all known measures of incremental confirmation fail to meet it. I describe this additional condition as well as a measure that meets it. The paper then applies the measure to the conjunction fallacy and proposes an explanation of the fallacy.

## 1. Justification and Confidence

This paper examines the degree of epistemic justification (hereafter simply "degree of justification") for accepting or rejecting propositions from the perspective of the dual goal of cognition, which is to increase true beliefs and reduce false beliefs. To be a little more precise, when we add propositions to our body of beliefs, the dual goal is to increase true beliefs but not to increase false beliefs. When we remove propositions from our body of beliefs, the dual goal is to reduce false beliefs but not to reduce true beliefs. Whether we are adding or removing propositions, the goal must have two components for obvious reasons. It is easy to increase true beliefs: Believe everything you can think of, including negations of what you already believe, and never abandon any beliefs. But, of course, we end up with numerous false beliefs, which is unacceptable. It is also easy to reduce false beliefs: Abandon all beliefs and don't form any new beliefs. But then we end up with no true beliefs, which is also unacceptable. The challenge is to balance the two demands. I will focus on cases of adding propositions to our body of beliefs, which is more straightforward than removing propositions from a tangled web of existing beliefs. The relevant goal of cognition is then to increase true beliefs but not to increase false beliefs. In this section I argue that when we understand epistemic justification from this perspective, we must reject the common view that the degree of justification for accepting a proposition is its probability.

To express the common view a little more precisely, the degree of justification for accepting the proposition $h$ given the evidence $e$ (based on the background knowledge $b$—this is suppressed in the following discussion) is the conditional probability of $h$ given $e$, or $P(h|e)$. It may seem that this view can take account of the dual goal of cognition. If we only care about increasing true beliefs, we set the probabilistic threshold of justification at the lowest possible level, viz. we are justified in accepting $h$ if and only if $P(h|e) \geq 0$, and accept any propositions we can think of. If we only care about not increasing false beliefs, we set the threshold degree at the highest possible level, viz. $P(h|e) \geq 1$, and reject all but absolutely certain propositions. Since neither approach serves the dual goal of cognition well, we set the threshold $t$ somewhere in between, depending on our degree of risk aversion—perhaps in consideration of the pragmatic context. However, it is well known that this view is in conflict with an intuitive principle about conjunction, viz., if we are justified in accepting each conjunct, then we are justified in accepting their conjunction. The conflict arises because it is possible for any non-trivial probabilistic threshold $t$ (i.e. $t \neq 0, 1$) that $P(h_1|e) \geq t, \ldots, P(h_n|e) \geq t$ but $P(h_1 \wedge \ldots \wedge h_n|e) < t$. When this happens, and if we apply the common view of justification, we are justified in accepting each of the propositions $h_1, \ldots, h_n$ but not their conjunction $h_1 \wedge \ldots \wedge h_n$. This is a violation of the intuitive principle. The lottery paradox (Kyburg 1961) and the preface paradox (Makinson 1965) are good illustrations of the difficulty, but I want to present the problem in a different way to focus on what I take to be the core issue.

Consider the set $H = \{h_1, \ldots, h_n\}$ of *probabilistically independent* propositions. To put it informally, these propositions have nothing to do with each other. The proposition $h_1$ could be about the demise of the Roman Empire, while the proposition $h_2$ could be about the salmon's immune system, and so forth. Let's assume that they remain probabilistically independent given the body of all available evidence $e$ that is relevant to these propositions.[1] To exclude trivial cases, we also assume that none of the propositions we consider is completely verified or completely refuted by the evidence $e$, i.e. $0 < P(h_1|e), \ldots, P(h_n|e) < 1$. Now, given their mutual irrelevance, one would expect that provided we are justified in accepting each of them, we are justified in accepting all of them. Here is the reasoning. First, we consider $h_1$ alone. We are justified in accepting $h_1$ because we are justified in accepting each of the propositions. Next we consider $h_2$. Since $h_1$ and $h_2$ are mutually irrelevant, we can evaluate $h_2$ independently of our acceptance of $h_1$. So, we are justified in accepting $h_2$ because we are justified in accepting each of the propositions. The same reasoning applies to $h_3$, $h_4$, and so on. As a result, we are justified in accepting all the propositions $h_1, \ldots, h_n$.

Notice that one forceful response to the lottery paradox does not apply to the present case. When the individually acceptable propositions are jointly inconsistent as in the lottery paradox, it could be plausibly argued that we can accept each of $h_1, \ldots, h_n$ by

---

[1] The evidence $e$ consists of $e_1, \ldots, e_n$ that respectively support $h_1, \ldots, h_n$. The propositions $h_1, \ldots, h_n$ are still probabilistically independent on condition of $e$ provided $e_1, \ldots, e_n$ are probabilistically independent. Note that $h_1, \ldots, h_n$ are *not* probabilistically independent on condition of $\neg e$ even if $e_1, \ldots, e_n$ are probabilistically independent, so the Fork Theorem (Reichenbach 1956, Section 19) does not apply to the present case.

itself, but not all of them together, i.e. we cannot accept $H = \{h_1, \ldots, h_n\}$ as a set. But this suggestion makes sense only if there is inconsistency—or at least some tension—among the propositions $h_1, \ldots, h_n$ while we are assuming in the present case that the propositions involved are mutually irrelevant. This allows us to evaluate each proposition independently even if we have already accepted some of the propositions, so that if we are justified in accepting each member, we are justified in accepting the set.

Some people may question the final move from the acceptance of the set $H = \{h_1, \ldots, h_n\}$ to the acceptance of the conjunction $h_1 \wedge \ldots \wedge h_n$ because $P(h_1 \wedge \ldots \wedge h_n|e)$ can be extremely low when the number of the conjuncts is large. How can we be justified in accepting a proposition that is almost certainly false? My response is twofold. First, there is no difference between accepting all of $h_1, \ldots, h_n$ *together* and accepting their conjunction $h_1 \wedge \ldots \wedge h_n$. Once we accept all the conjuncts together, it is unreasonable not to accept their conjunction. Second, we should distinguish the degree of justification from the degree of confidence. The subject of this paper is the degree of justification motivated by the dual goal of cognition, which is to increase true beliefs and reduce false beliefs. The degree of *confidence* serves other purposes, most notably the calculation of the expected utility. In order to play that role, the degree of confidence should be proportional to the probability. So, we should be less *confident* in the conjunction $h_1 \wedge \ldots \wedge h_n$ than we are in any of the conjuncts $h_1, \ldots, h_n$. However, that does not mean that we are less justified in accepting the conjunction than we are in accepting any conjunct. Accepting the conjunction is riskier than accepting a conjunct because the conjunction has a lower probability than any conjunct. But this higher risk is counterbalanced by the greater potential gain in true beliefs. From the perspective of the dual goal of cognition, the risk of adding false beliefs is not the sole determinant of the degree of justification—the potential gain in true beliefs is also a factor. Once we distinguish the degree of justification from the degree of confidence in this way, the common view that the degree of justification is the conditional probability of the proposition given the evidence loses its appeal. Even if the conditional probability is low, we may still be justified in accepting the proposition if the potential gain in truth is sufficiently high.

## 2. Formalizing the Risk and the Potential Gain

This section formalizes the two factors that affect the degree of justification—the risk of adding false beliefs and the potential gain in truth beliefs—in probabilistic terms.[2] First, the risk of increasing false beliefs is inversely related to the conditional probability of the proposition given the evidence. The higher the evidence makes the probability of the proposition, the lower the risk of increasing false beliefs. Since the risk of increasing false beliefs is inversely related to the degree of justification, the conditional probability of the proposition given the evidence is directly (positively) related to the degree of justification. There is no surprise here. The other factor, the potential gain in true beliefs, may seem less clear. Obviously, we cannot simply count the *number* of potentially true beliefs. Adding the set $H = \{h_1, \ldots, h_n\}$ to our body of beliefs is no different from adding

[2] See Huber (2008a; 2008b) for a similar two-factor approach to the formal assessment of scientific theories. Huber calls the two factors “plausibility” and “informativeness”.

the singleton $H^* = \{h_1 \wedge \ldots \wedge h_n\}$ though the former contains many more propositions, and hence many more potentially true beliefs. A more sensible approach is to measure the potential gain in true beliefs by the amount of information the proposition (or the conjunction of the propositions if a set of propositions is added) carries. Since the amount of information the proposition carries is inversely related to its *prior probability*, we can capture the potential gain in true beliefs in probabilistic terms.

We can see the inverse relation between the amount of information and the prior probability in two steps. First, the degree of *specificity* is directly (positively) related to the amount of information. The more specifically the proposition describes the world, the larger *amount of information* it carries. Second, the degree of specificity is inversely related to the *prior probability*. The more specifically the proposition describes the world, the lower its prior probability is. By combining these two steps, we see that the amount of information that the proposition carries is inversely related to its prior probability. Further, since the amount of information the proposition carries is directly (positively) related to the degree of justification, the prior probability of the proposition is inversely related to the degree of justification. To express this more intuitively, if the level of risk is the same (if the conditional probability of the proposition given the evidence is the same), a proposition that describes the world more specifically (and thus has a lower prior probability) is more worthy of adding to our body of beliefs because the per-unit-of-information risk is lower.

We put all these together to state that the degree of justification $J(h, e)$ for the proposition $h$ given the evidence $e$ is directly (positively) related to the conditional probability $P(h|e)$ and inversely related to the prior probability $P(h)$. Note that under this conception the degree of justification looks much like the degree of incremental confirmation (hereafter simply “degree of confirmation”). There have been many proposals in the literature to formally measure the degree of confirmation. Here I mention only two of them, the difference measure $C_D(h, e)$ and the ratio measure $C_R(h, e)$:

$$C_D(h, e) = P(h \mid e) - P(h)$$

$$C_R(h, e) = \frac{P(h \mid e)}{P(h)}$$

In both measures, the degree of confirmation is directly (positively) related to the conditional probability and inversely related to the prior probability, and that is the way it should be for any plausible measure of confirmation.

The question arises at this point whether the degree of justification is simply the degree of confirmation. The question has two parts: (1) whether an additional condition exists that the degree of confirmation should satisfy but the degree of justification need not, and (2) whether an additional condition exists that the degree of justification should satisfy but the degree of confirmation need not. The next section addresses these two questions.

## 3. Justification and Confirmation

There is a general consensus in the literature that in addition to being an increasing function of the conditional probability and a decreasing function of the prior probability, the degree of confirmation should have a constant neutral value $k$ when $P(h|e) = P(h)$, regardless of $P(h)$. The idea is that when the evidence $e$ has no impact on the proposition $h$ and thus $P(h|e) = P(h)$, the evidence neither confirms nor disconfirms the proposition. So, the degree of confirmation in such cases should be the same, regardless of the prior probability of the proposition. Let's call this requirement the equi-neutrality condition. The equi-neutrality condition is satisfied by all known measures of confirmation. For example, the condition $P(h|e) = P(h)$ makes the difference measure $C_D(h, e) = P(h|e) – P(h)$ constant at zero; it makes the ratio measure $C_R(h, e) = P(h|e) / P(h)$ constant at one. We can adjust any measure of confirmation to make the neutral value zero by subtracting the constant value $k$ from it.[3] For example, if we subtract one from the ratio measure, the new measure $C_R{}^*(h, e) = P(h|e) / P(h) – 1$ has its neutral value at zero. So, I will assume hereafter that the neutral degree of confirmation is zero, i.e. $C(h, e) = 0$ when $P(h|e) = P(h)$.

I want to argue in this section that the degree of *justification* should also satisfy the equi-neutrality condition—i.e. $J(h, e) = 0$ when $P(h|e) = P(h)$, regardless of $P(h)$. In other words, although there is an additional condition (beyond being an increasing function of the conditional probability and a decreasing function of the prior probability) that the degree of confirmation should satisfy, the degree of justification should also satisfy it. The basis of my argument for the equi-neutrality of justification is the case of conjunction mentioned in Section 1, namely: If the propositions $h_1, \ldots, h_n$ are probabilistically independent, both unconditionally and conditionally given the evidence $e$, and if each of them is justified by the evidence $e$ (with regard to some threshold degree $t$), then so is their conjunction $h_1 \wedge \ldots \wedge h_n$. The converse should also hold: If the propositions $h_1, \ldots, h_n$ are probabilistically independent, both unconditionally and conditionally given the evidence $e$, and if each of them is not justified by $e$ (with regard to some threshold degree $t$), then neither is their conjunction $h_1 \wedge \ldots \wedge h_n$. I call the combination of these two conditions the General Conjunction Requirement (GCR).

An immediate consequence of GCR is the following Special Conjunction Requirement (SCR): If the propositions $h_1, \ldots, h_n$ are probabilistically independent, both unconditionally and conditionally given the evidence $e$, and if each of them is justified to the same degree $j$, then so is their conjunction $h_1 \wedge \ldots \wedge h_n$. I show here that GCR entails SCR by proving its contraposition. Suppose measure $J(h, e)$ of justification fails to satisfy SCR, and thus for some evidence $e$ and some probabilistically independent (both unconditionally and conditionally given $e$) propositions $h_1, \ldots, h_n$, $J(h_1, e) = \ldots = J(h_n, e) = j$ but $J(h_1 \wedge \ldots \wedge h_n, e) = j + \alpha$ for some $\alpha \neq 0$. We can see that $J(h, e)$ violates GCR as follows. Set the threshold of justification at $j + \alpha/2$. If $\alpha < 0$, then each of $h_1, \ldots, h_n$ is justified by $e$, but their conjunction $h_1 \wedge \ldots \wedge h_n$ is not. If $\alpha > 0$, then each of $h_1, \ldots, h_n$ is not justified by $e$, but their conjunction $h_1 \wedge \ldots \wedge h_n$ is. Either way, GCR is violated. So, GCR entails SCR. Further, if we assume that $J(h, e)$, which is of the form $F(P(h|e), P(h))$, is a continuous function, then SCR entails equi-neutrality (see Appendix 1 for proof).

[3] The obtained measure $C_X{}^*(h, e) = C_X(h, e) – k$ is *ordinally equivalent* to the original measure $C_X(h, e)$, i.e. for any two pairs $< h_1, e_1>$ and $< h_2, e_2>$, $C_X{}^*(h_1, e_1)$ >/=/< $C_X{}^*(h_2, e_2)$ if and only if $C_X(h_1, e_1)$ >/=/< $C_X(h_2, e_2)$, respectively. For many purposes, ordinally equivalent measures are essentially the same measure.

Putting all these together, we conclude that $J(h, e)$ should satisfy the equi-neutrality condition since $J(h, e)$ should satisfy GCR, which entails SCR, which in turn entails equi-neutrality.

This result may look suspect. When the evidence affects neither $h_1$ nor $h_2$, are we no more justified in accepting $h_1$ than in accepting $h_2$ even if $h_1$ is almost certainly true while $h_2$ is almost certainly false? My response is again the distinction between the degree of confidence and the degree of justification. We should certainly have more confidence in $h_1$ than in $h_2$ when $P(h_1)$ is higher than $P(h_2)$, but it does not follow that we are more justified in accepting $h_1$ than we are in accepting $h_2$. Though $h_2$ is more likely to be false than $h_1$ is, the higher risk is offset by the greater potential gain we make if $h_2$ turns out to be true because $h_2$, whose prior probability is lower, carries more information than $h_1$ does. So, if the degree of justification is to serve the dual goal of cognition, it is not unreasonable to assign the same degree of justification to $h_1$ and $h_2$.[4]

To summarize what we have uncovered so far, the degree of justification for the proposition $h$ given the evidence $e$ is directly (positively) related to its conditional probability $P(h|e)$ and inversely related to the prior probability $P(h)$. Further, the degree of justification should also satisfy the equi-neutrality condition, i.e. $J(h, e) = 0$ when $P(h|e) = P(h)$, regardless of $P(h)$. Since these are the standard requirements for a measure of *confirmation*, a measure of justification is also a measure of confirmation. However, the converse is not true. Not all plausible measures of confirmation can serve as a measure of justification because the latter must satisfy the General Conjunction Requirement (GCR), while there is no reason to require that a measure of confirmation should satisfy GCR. Indeed none of the many measures of confirmation proposed in the literature satisfies GCR.[5] So, none of them is a measure of justification. We need to formulate a new measure of confirmation that meets GCR.

## 4. Formal Measure of Justification

This section describes a formal measure $J(h, e)$ of justification for the proposition $h$ given the evidence $e$. In order to facilitate the task, I want to describe one further consequence of GCR. We saw in Section 3 that the degree of justification should be equi-neutral. It turns out that the degree of justification should also be *equi-maximal*. It is obvious already that for any given $P(h)$, $J(h, e)$ should be the highest when $P(h|e) = 1$ because $J(h, e)$ is an increasing function of $P(h|e)$. Equi-maximality requires further that this highest value should be constant, regardless of $P(h)$. Intuitively, this means that when the evidence $e$ makes the proposition $h$ certain, we are justified in accepting $h$ to the highest

[4] Equi-neutrality of epistemic justification explains the intuition that in the absence of some inside information we cannot assert that a given lottery ticket does not win even if the probability for that proposition is extremely high (Williamson 2000, p. 246). The reason is that when there is no relevant evidence beyond the background knowledge, there is no positive justification at all for the proposition, no matter how high its prior probability is.

[5] See Fitelson (1999, 2001), Crupi, Tentori, and Gonzalez (2007) for the growing list of confirmation measures.

possible degree, regardless of the prior probability of $h$.[6] This is a sensible thing to say about the degree of justification, but it is also a consequence of SCR (see Appendix 2 for proof) and hence of GCR.

Let us see what $J(h, e)$ should look like in light of the requirements we have uncovered. First, $J(h, e)$ should be an increasing function of $P(h|e)$ and a decreasing function of $P(h)$. There are two natural ways for $J(h, e)$ to meet these requirements, namely, the difference-based measures $J_D$ and the ratio-based measures $J_R$:

$$J_D(h, e) = f(P(h \mid e)) - g(P(h))$$
$$J_R(h, e) = \frac{f(P(h \mid e)}{g(P(h))}$$

where both $f$ and $g$ are increasing functions. The second set of requirements is equi-neutrality and equi-maximality. If we set the neutral value at zero and the maximum value at one, then:

$$J(h, e) = 0 \text{ when } P(h|e) = P(h)$$
$$J(h, e) = 1 \text{ when } P(h|e) = 1$$

We need to adjust the difference-based measures $J_D(h, e)$ and the ratio-based measures $J_R(h, e)$ to meet this second set of requirements.

We start with the difference-based measures. When $P(h|e) = P(h)$, $J_D(h, e) = f(P(h)) - g(P(h))$. Since this value should be zero regardless of $P(h)$, $f$ and $g$ should be the same function. This means that $J_{D*}(h, e) = f(P(h|e)) - f(P(h))$. Further, when $P(h|e) = 1$, $J_D*(h, e) = f(1) - f(P(h))$. Since this value should be one regardless of $P(h)$, we need to "normalize" $J_D*(h, e)$ by dividing it by $f(1) - f(P(h))$, to obtain $J_D**(h, e) = [f(P(h|e) - f(P(h))] / [f(1) - f(P(h))]$. This measure satisfies both the first and second sets of requirements. We turn next to the ratio-based measures $J_R(h, e)$. When $P(h|e) = P(h)$, $J_R(h, e) = f(P(h)) / g(P(h))$. Since this value should be zero regardless of $P(h)$, we subtract $f(P(h)) / g(P(h))$ from $J_R(h, e)$ to obtain $J_R*(h, e) = [f(P(h|e)) / g(P(h))] - [f(P(h)) / g(P(h))] = [f(P(h|e)) - f(P(h))] / g(P(h))$. Further, when $P(h|e) = 1$, $J_R*(h, e) = [f(1) - f(P(h))] / g(P(h))$. Since this value should be one regardless of $P(h)$, $g(P(h))$ should be $f(1) - f(P(h))$, so that $J_R**(h, e) = [f(P(h|e)) - f(P(h))] / [f(1) - f(P(h))]$. This turns out to be the same as $J_D**(h, e)$. So, whether we start from the difference-based measures $J_D(h, e)$ or the ratio-based measures $J_R(h, e)$, we arrive at the same general formula, $J_G(h, e) = [f(P(h|e)) - f(P(h))] / [f(1) - f(P(h))]$, to satisfy the second sets of requirements. The remaining task is to determine the function $f$, so that $J_G(h, e)$ satisfies the General Conjunction Requirement.

This task is not trivial. If we take $f$ to be the identity function, $f(x) = x$, then the degree of justification will be $J_G*(h, e) = [P(h|e) - P(h)] / [1 - P(h)]$.[7] But $J_G*(h, e)$ fails to meet the SCR (and hence GCR) even for $n = 2$, i.e. even when the conjunction has

---

[6] We assume that the proposition $h$ is not already certain, so it cannot the case that $P(h|e) = P(h) = 1$ to make $J(h, e)$ both neutral and maximal.

[7] This is the positive half of Crupi, Tentori, and Gonzalez's (2007) measure $Z$ of confirmation.

only two conjuncts (proof omitted). The problem is solved by making $f$ a logarithmic function. If we choose 2 as the base of logarithm, as it is common in measuring the amount of information, then we obtain the following measure $J_G{}^{**}(h, e)$:[8]

$$J_G{}^{**}(h, e) = \frac{\log_2 P(h \mid e) - \log_2 P(h)}{\log_2 1 - \log_2 P(h)}$$
$$= \frac{\log_2 P(h \mid e) - \log_2 P(h)}{-\log_2 P(h)}$$

$J_G{}^{**}(h, e)$ meets GCR (see Appendix 3 for proof), so it is a measure of justification. From now on, I will write $J_G{}^{**}(h, e)$ simply as $J(h, e)$.

Once we find a measure of justification, the next natural question is whether it is the only measure of justification. It turned out that there are many others, i.e. we can construct many measures of confirmation that satisfy GCR and thus can serve as measures of justification. Some of them differ from $J(h, e)$ in an interesting way. For example, $J(h, e)$ has the infinite range $(-\infty, 1]$, while Atkinson's (2009) measure $J'(h, e)$ has the finite range $[-1, 1]$. However, it also turned out that all measures of justification are ordinally equivalent to each other, and thus to $J(h, e)$.[9] In other words, $J(h, e)$ is the unique measure of justification, up to ordinal equivalence.

Two more remarks are in order. First, $J(h, e)$ is related to the log ratio measure of confirmation, which is $C_{LR}(h, e) = \log_2 [P(h|e) / P(h)]$ if we choose 2 as the base of logarithm. Note that the numerator of $J(h, e)$ is the log ratio measure of confirmation, i.e. $\log_2 P(h|e) - \log_2 P(h) = \log_2 [P(h|e) / P(h)]$. The denominator of $J(h, e)$ is the highest value of $C_{LR}(h, e)$ reached when $P(h|e) = 1$, i.e. $-\log_2 P(h) = \log_2 [1 / P(h)]$. This means that $J(h, e)$ is the "normalized" log ratio measure of confirmation.[10]

Second, $J(h, e)$ has a simple and intuitive meaning when we express it in the language of information. According to the standard mathematical theory of information, the amount of information that $h$ carries is $I(h) = -\log_2 P(h)$. The rationale for this measure is easy to see by an example. If the probability of the proposition $h$ is 1/8, then the amount of information that $h$ carries is $I(h) = -\log_2 1/8 = -\log_2 2^{-3} = 3$. This means that knowing $h$ with certainty gives us 3 bits of information. $I(h)$ is commonly referred to as "self information" because it is the amount of information on $h$ that we gain when it becomes certain that $h$ is true. Meanwhile, the amount of information on $h$ that we gain when it becomes certain that $e$ is true is called "mutual information" and is defined as follows: $I(h, e) = \log_2 P(h|e) - \log_2 P(h)$.[11] To see its intuitive meaning, suppose the prior

---

[8] The assumption that $P(h) \neq 1$ (see note 6 above) ensures that the denominator $-\log_2 P(h)$ is not zero.

[9] See Appendix 4 for proof. Atkinson (2009) obtained the same result independently with an illuminating alternative proof.

[10] Crupi, Tentori, and Gonzalez (2007) point out that many known measures of confirmation become ordinally equivalent to their preferred measure $Z$ (see note 7 above) when they are "normalized," but the log ratio measure is not one of them.

[11] $I(x, y)$ is called "mutual" information because it follows from the definition that $I(x, y) = I(y, x)$.

probability of the proposition $P(h)$ is 1/8 and the evidence $e$ raises its probability to $P(h|e)$ = 1/2. Then, the amount of mutual information is $I(h, e) = \log_2 2^{-1} - \log_2 2^{-3} = 2$. This means that we gain 2 bits of information on $h$ when we obtain the evidence $e$. The point to note is that the numerator of $J(h, e)$ is the mutual information $I(h, e) = \log_2 P(h|e) - \log_2 P(h)$, while the denominator of $J(h, e)$ is the self information $I(h) = -\log_2 P(h)$. So, $J(h, e)$ turns out to be the ratio of the mutual information to the self information:

$$J(h, e) = \frac{I(h,e)}{I(h)}$$

This expression allows us to interpret $J(h, e)$ as the degree of justification in a natural sense. Self information $I(h)$ is the amount of information we *register* when we add $h$ to our body of beliefs. I want to call it "registered information." Meanwhile mutual information $I(h, e)$ is the amount of information on $h$ we gain from the evidence $e$. So, I call it "earned information." If we use this terminology, the degree of justification $J(h, e)$ is the ratio of the earned information to the registered information. The higher the ratio is, the more justified we are in accepting (registering) the proposition. This makes good sense if the degree of justification is to serve the dual goal of cognition, which is to increase true beliefs and reduce false beliefs.

## 5. The Conjunction Fallacy

This section applies the measure of justification $J(h, e)$ to the analysis of the conjunction fallacy. The conjunction fallacy is the fallacy of assigning a higher probability to a conjunction $h_1 \wedge h_2$ than to its conjunct $h_1$ (or $h_2$). Since the conjunction $h_1 \wedge h_2$ logically entails the conjunct $h_1$, the conjunction cannot have a higher probability than the conjunct, but it is well known that people are prone to commit this fallacy in certain contexts. The most famous is the Linda problem (Tversky and Kahneman 1983), in which the two conjuncts are:

> $h_1$: Linda is a bank teller.
> $h_2$: Linda is active in the feminist movement.

The participants in the experiment receive the following information:

> $e$: Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Upon receiving this information, a large majority of the participants answer that $h_1 \wedge h_2$, is more probable than $h_1$, committing the conjunction fallacy.

Tversky and Kahneman explain the fallacy by the representativeness heuristic, i.e. given $e$, most participants judge that Linda is more representative of a feminist banker than of a banker, and they solely rely on this judgment in assigning a higher probability to $h_1 \wedge h_2$ than to $h_1$. More formal analyses are also possible. Shafir, Smith, and Osherson

(1990) propose that most participants focus on *likelihood*, i.e. the conditional probability of the evidence given the hypothesis. According to this analysis, most participants compare the two likelihoods $P(e|h_1 \wedge h_2)$ and $P(e|h_1)$, instead of comparing the two conditional probabilities $P(h_1 \wedge h_2|e)$ and $P(h_1|e)$ as they should. Another possibility is that most participants focus on the degree of coherence between the evidence and the hypothesis. We can make it a formal analysis by plugging in any of the many probabilistic measures of coherence available in the literature.[12] Yet another possibility is that most participants focus on the degree of confirmation (Sides *et al*. 2002), i.e. they compare the degrees to which the evidence raises the probabilities of the two hypotheses, $h_1$ and $h_1 \wedge h_2$. In support of this idea Crupi, Fitelson, and Tentori (2008) show that the confirmation analysis is *robust*. That is to say, in Linda-like cases—which they characterize by the two conditions (1) $P(h_2|e \wedge h_1) > P(h_2|h_1)$ and (2) $P(h_1|e) < P(h_1)$—the evidence $e$ confirms the conjunction $h_1 \wedge h_2$ more than it does the conjunct $h_1$ by *any measure of confirmation* that has been proposed in the literature.

These analyses offer competing accounts of the cognitive process responsible for the fallacy, or which features of the case most participants focus on. But I am more interested in the conditions under which the fallacy is common than in the cognitive process. To use Marr's (1982) distinction, I am more interested in the *computation* (the input-output relation) that is accomplished than in the *algorithm* for the computation. The three formal analyses—by likelihood, by coherence, and by confirmation—are similar at the computational level. In fact they are formally equivalent if we determine the degree of coherence by Shogenji's (1999) measure $S(x_1, \ldots, x_n) = P(x_1 \wedge \ldots \wedge x_n) / [P(x_1) \times \ldots \times P(x_n)]$ and the degree of confirmation by the ratio measure $C_R(h, e) = P(h \wedge e) / P(h)$.[13] I have no reason to think these formal conditions are seriously at odds with empirical data, but I still propose my own analysis. The reason for the proposal is not a better fit with the empirical data but a better explanation of *why* the fallacy occurs.

Here is my proposal (the justification analysis): The conjunction fallacy is common when the degree of justification for the conjunction is higher than the degree of justification for the conjunct, i.e. $J(h_1 \wedge h_2, e) > J(h_1, e)$. Since $J(h, e)$ is also a measure of confirmation, the proposal is a variant of the confirmation analysis. Given the robustness of the confirmation analysis, it is not surprising that the justification analysis gives the right prediction in Linda-like cases, i.e. when (1) $P(h_2|e \wedge h_1) > P(h_2|h_1)$ and (2) $P(h_1|e) < P(h_1)$, the evidence justifies the conjunction more than it does the conjunct, or $J(h_1 \wedge h_2, e) > J(h_1, e)$ (see Appendix 5 for proof), so that the conjunction fallacy should be common in Linda-like cases. The attraction of the justification analysis is its explanation of why the fallacy occurs, viz., the fallacy occurs because we tend to utilize cognitive processes appropriate for choosing better justified propositions, even when that is not our task. These justification-oriented processes serve the dual goal of cognition well, so their persistent use is generally a good epistemic policy. However, it causes trouble in cases where our task is not to choose better justified propositions but to choose more *probable* propositions. The explanation makes the conjunction fallacy understandable.

---

[12] See Meijs (2005) for a survey.

[13] That is to say, $P(e|h_1 \wedge h_2) > P(e|h_1)$ iff $S(h_1 \wedge h_2, e) > S(h_1, e)$ iff $C_R(h_1 \wedge h_2, e) > C_R(h_1, e)$ (proof omitted).

The justification analysis is compatible with different theories of the cognitive process. One of the cognitive processes mentioned above may be responsible for the conjunction fallacy. If so, my proposal is that we utilize that cognitive process, not because it guides us to choose propositions with high degrees of representativeness, likelihood, coherence, or confirmation *per se*, but because it guides us to choose propositions with high degrees of justification. In other words, the computational objective of the process is to choose better justified propositions.

I want to note that although I want to make the conjunction fallacy understandable, I do not subscribe to the view that the conjunction fallacy (or the "conjunction effect") can be explained by semantic variance (cf. Hertwig and Gigerenzer 1999). For example, I do not think that many people interpret the word "probable" to mean *justified* and that their judgment is correct under this interpretation. The betting case provides strong evidence against the semantic account. It is known that the conjunction fallacy occurs even in betting, e.g. many people are more willing to bet on $h_1 \wedge h_2$ than on $h_1$ in the Linda case for the same reward (Tversky and Kahneman 1983, p. 300). There is no semantic excuse for this behavior since the situation itself requires the assessment of probabilities. Some people question the reality of the conjunction fallacy on other grounds. It has been reported that changing the problem structure—e.g. expressing the problem in terms of frequencies instead of probabilities—reduces the occurrence of cognitive fallacies, including the conjunction fallacy (Gigerenzer 1991). But if the justification analysis is correct, the fallacy should be less frequent in those contexts where people are less accustomed to choosing better justified propositions automatically. If this is born out, the reduction of the fallacy in such contexts strengthens the case for the justification analysis.

## 6. Conclusion

When we aim at the dual goal of cognition, the degree of justification for accepting the proposition should not be its conditional probability given the evidence, as it is commonly thought. We have compelling reason to adopt $J(h, e)$ as our formal measure of justification. It has a simple and intuitive meaning as the ratio of the earned information to the registered information, and it is the only measure (up to ordinal equivalence) that meets the General Conjunction Requirement. I already mentioned its relevance to the lottery paradox and the preface paradox in Section 1, and showed how it helps the analysis of the conjunction fallacy in Section 5. Another significant area of application is logical closure of knowledge. Even if $p$ logically entails $q$, the degree of justification for $q$ can be lower than that for $p$, as the conjunction fallacy exemplifies. This means that knowledge is not closed under (known) logical entailment if a certain degree of justification is a necessary condition for knowledge. I suspect that we need to reconsider many issues of cognitive science and normative epistemology in light of the new understanding of epistemic justification.[14]

[14] An earlier version of this paper was presented at the workshop *Probability, Confirmation and Fallacies* in Leuven, Belgium, in April 2008. I would like to thank its organizers Jeanne Peijnenburg, David Atkinson and Igor Douven, and the participants of the workshop for many stimulating discussions. Special thanks to David Atkinson and Branden Fitelson for valuable post-conference correspondence.

# Appendices

## 1. Equi-Neutrality

Suppose $J(h, e)$, which is of the form $F(P(h|e), P(h))$, is a continuous function, and $J(h, e)$ satisfies SCR. Then, for any two pairs $<h_i, e_i>$ and $<h_j, e_j>$, if $P(h_i|e_i) = P(h_i)$ and $P(h_j|e_j) = P(h_j)$, then $J(h_i, e_i) = J(h_j, e_j)$.

**Proof:**

Let $\log_{P(h_i)} P(h_j) = r$, so that $[P(h_i)]^r = P(h_j)$. $r > 0$ because $0 < P(h_i), P(h_j) < 1$. Since $J(h, e)$ is a continuous function, it suffices to show that the claim holds for any two pairs $<h_i, e_i>$ and $<h_j, e_j>$ such that $[P(h_i)]^q = P(h_j)$ where $q$ is a positive rational number. Let $<m, n>$ be the smallest pair of positive integers such that $n/m = q$, so that $[P(h_i)]^n = [P(h_j)]^m$. Choose probabilistically independent (both unconditionally and conditionally on $e_i$) propositions $h_1, \ldots, h_n$, and probabilistically independent (both unconditionally and conditionally on $e_j$) propositions $h_{n+1}, \ldots, h_{n+m}$ such that:[15]

(i) $[P(h_i)]^n = [P(h_j)]^m$
(ii) $P(h_i) = P(h_1) = \ldots = P(h_n)$
(iii) $P(h_j) = P(h_{n+1}) = \ldots = P(h_{n+m})$
(iv) $P(h_i|e_i) = P(h_1|e_i) = \ldots = P(h_n|e_i)$
(v) $P(h_j|e_j) = \mathrm{P}(h_{n+1}|e_j) = \ldots = \mathrm{P}(h_{n+m}|e_j)$

It follows from (ii) and (iv) that $J(h_i, e_i) = J(h_1, e_i) = \ldots = J(h_n, e_i)$. So, by SCR:

$$J(h_i, e_i) = J(h_1 \wedge \ldots \wedge h_n, e_i) \tag{1}$$

Similarly, it follows from (iii) and (v) that $J(h_j, e_j) = J(h_{n+1}, e_j) = \ldots = J(h_{n+m}, e_j)$. So, by SCR:

$$J(h_j, e_j) = J(h_{n+1} \wedge \ldots \wedge h_{n+m}, e_j) \tag{2}$$

---

[15] For example, think of $n$ urns of colored marbles, for each of which the probability of drawing a red marble is the same as $P(h_i)$, and $m$ urns of colored marbles, for each of which the probability of drawing a red marble is the same as $P(h_j)$. To satisfy the conditions $P(h_i|e_i) = P(h_i)$ and $P(h_j|e_j) = P(h_j)$ of the theorem (in addition to (i) through (v)), the $n$ urns must have nothing to do with $e_i$ and the $m$ urns must have nothing to do with $e_j$.

Since $h_1, \ldots, h_n$ are probabilistically independent, $P(h_1 \wedge \ldots \wedge h_n) = [P(h_i)]^n$ from (ii). Similarly, since $h_{n+1}, \ldots, h_{n+m}$ are probabilistically independent, $P(h_{n+1} \wedge \ldots \wedge h_{n+m}) = [P(h_j)]^m$ from (iii). So, it follows from (i) that:

$$P(h_1 \wedge \ldots \wedge h_n) = P(h_{n+1} \wedge \ldots \wedge h_{n+m}) \qquad (3)$$

Since $h_1, \ldots, h_n$ are probabilistically independent conditionally on $e_i$, $P(h_1 \wedge \ldots \wedge h_n|e_i) = [P(h_i|e_i)]^n = [P(h_i)]^n$ from (iv) and from the condition $P(h_i|e_i) = P(h_i)$ of the theorem. Similarly, since $h_{n+1}, \ldots, h_{n+m}$ are probabilistically independent conditionally on $e_j$, $P(h_{n+1} \wedge \ldots \wedge h_{n+m}|e_j) = [P(h_j|e_j)]^m = [P(h_j)]^m$ from (v) and from the condition $P(h_j|e_j) = P(h_j)$ of the theorem. So, it follows from (i) that:

$$P(h_1 \wedge \ldots \wedge h_n|e_i) = P(h_{n+1} \wedge \ldots \wedge h_{n+m}|e_j) \qquad (4)$$

From (3) and (4) it follows that:

$$J(h_1 \wedge \ldots \wedge h_n, e_i) = J(h_{n+1} \wedge \ldots \wedge h_{n+m}, e_j) \qquad (5)$$

From (1), (2) and (5) it follows that $J(h_i, e_i) = J(h_j, e_j)$. ■

## 2. Equi-Maximality

Suppose $J(h, e)$ is a justification measure (i.e. a confirmation measure that satisfies GCR and hence SCR). Then, for any two pairs $<h_i, e_i>$ and $<h_j, e_j>$, if $P(h_i|e_i) = P(h_j|e_j) = 1$, then $J(h_i, e_i) = J(h_j, e_j)$.

**Proof:**

Assume without loss of generality that $P(h_i) \leq P(h_j)$. It follows from the condition $P(h_i|e_i) = P(h_j|e_j)$ of the theorem that:

$$J(h_i, e_i) \geq J(h_j, e_j) \qquad (1)$$

since the confirmation measure $J(h, e) = F(P(h|e), P(h))$ is a decreasing function of $P(h)$. Choose probabilistically independent propositions $h_1, \ldots, h_n$ such that:

(i) $[P(h_j)]^n \leq P(h_i)$
(ii) $P(h_j) = P(h_1) = \ldots = P(h_n)$

It follows from (ii) and $P(h_1|h_1 \wedge \ldots \wedge h_n) = \ldots = P(h_n|h_1 \wedge \ldots \wedge h_n) = 1$ that $J(h_1, h_1 \wedge \ldots \wedge h_n) = \ldots = J(h_n, h_1 \wedge \ldots \wedge h_n)$. But $h_1, \ldots, h_n$ are probabilistically independent, and $h_1, \ldots, h_n$ are also trivially probabilistically independent conditionally on $h_1 \wedge \ldots \wedge h_n$ because $P(h_1|h_1 \wedge \ldots \wedge h_n) = \ldots = P(h_n|h_1 \wedge \ldots \wedge h_n) = 1$. So, by SCR:

$$J(h_1, h_1 \wedge \ldots \wedge h_n) = J(h_1 \wedge \ldots \wedge h_n, h_1 \wedge \ldots \wedge h_n) \qquad (2)$$

Also, it follows from the condition $P(h_j|e_j) = 1$ of the theorem and from $P(h_1|h_1 \wedge \ldots \wedge h_n) = 1$ that $P(h_j|e_j) = P(h_1|h_1 \wedge \ldots \wedge h_n)$. Further, $P(h_j) = P(h_1)$ from (ii). So,

$$J(h_j, e_j) = J(h_1, h_1 \wedge \ldots \wedge h_n) \qquad (3)$$

It follows from (2) and (3) that:

$$J(h_j, e_j) = J(h_1 \wedge \ldots \wedge h_n, h_1 \wedge \ldots \wedge h_n) \qquad (4)$$

Meanwhile, it follows from (ii) that $P(h_1 \wedge \ldots \wedge h_n) = [P(h_j)]^n$ since $h_1, \ldots, h_n$ are probabilistically independent. But $[\mathrm{P}(h_\mathrm{j})]^n \leq \mathrm{P}(h_\mathrm{i})$ from (i). So,

$$P(h_1 \wedge \ldots \wedge h_n) \leq P(h_i) \qquad (5)$$

while it follows from the condition $P(h_i|e_i) = 1$ of the theorem and from $P(h_1 \wedge \ldots \wedge h_n|h_1 \wedge \ldots \wedge h_n) = 1$ that:

$$P(h_1 \wedge \ldots \wedge h_n|h_1 \wedge \ldots \wedge h_n) = P(h_i|e_i) \qquad (6)$$

It follows from (5) and (6) that:

$$J(h_1 \wedge \ldots \wedge h_n, h_1 \wedge \ldots \wedge h_n) \geq J(h_i, e_i) \qquad (7)$$

since the confirmation measure $J(h, e)$ is a decreasing function of $P(h)$. It follows from (4) and (7) that:

$$J(h_j, e_j) \geq J(h_i, e_i) \qquad (8)$$

It follows from (1) and (8) that $J(h_i, e_i) = J(h_j, e_j)$. ■

## 3. General Conjunction Requirement

Suppose $h_1, \ldots, h_n$ are probabilistically independent (both unconditionally and conditionally on $e$) and $P(h_1), \ldots, P(h_n) < 1$. Then, (i) if $J(h_1, e), \ldots, J(h_n, e) \geq t$, then $J(h_1 \wedge \ldots \wedge h_n, e) \geq t$; (ii) if $J(h_1, e), \ldots, J(h_n, e) < t$, then $J(h_1 \wedge \ldots \wedge h_n, e) < t$.

**Proof:**

$$J(h_1 \wedge \ldots \wedge h_n, e) = \frac{\log_2 P(h_1 \wedge \ldots \wedge h_n \mid e) - \log_2 P(h_1 \wedge \ldots \wedge h_n)}{-\log_2 P(h_1 \wedge \ldots \wedge h_n)}$$

$$= \frac{\log_2 \prod_{i=1}^{n} P(h_i \mid e) - \log_2 \prod_{i=1}^{n} P(h_i)}{-\log_2 \prod_{i=1}^{n} P(h_i)} \qquad \text{[from independence]}$$

$$= \frac{\sum_{i=1}^{n} \log_2 P(h_i \mid e) - \sum_{i=1}^{n} \log_2 P(h_i)}{-\sum_{i=1}^{n} \log_2 P(h_i)}$$

$$= \frac{\sum_{i=1}^{n} [\log_2 P(h_i \mid e) - \log_2 P(h_i)]}{\sum_{i=1}^{n} -\log_2 P(h_i)} \qquad (1)$$

(i) Suppose $J(h_1, e), \ldots, J(h_n, e) \geq t$. Then, for $i = 1, \ldots, n$, there is some $\alpha_i \geq 0$ such that:

$$J(h_i, e) = \frac{\log_2 P(h_i \mid e) - \log_2 P(h_i)}{-\log_2 P(h_i)}$$
$$= t + \alpha_i$$

So,

$$\log_2 P(h_i|e) - \log_2 P(h_i) = (t + \alpha_i)\,[-\log_2 P(h_i)] \qquad (2)$$

By plugging (2) into (1) above, we obtain:

$$J(h_1 \wedge \ldots \wedge h_n, e) = \frac{\sum_{i=1}^{n} (t+\alpha_i)[-\log_2 P(h_i)]}{\sum_{i=1}^{n} -\log_2 P(h_i)}$$

$$= \frac{t\sum_{i=1}^{n} -\log_2 P(h_i) + \sum_{i=1}^{n} \alpha_i[-\log_2 P(h_i)]}{\sum_{i=1}^{n} -\log_2 P(h_i)}$$

$$= t + \frac{\sum_{i=1}^{n} \alpha_i[-\log_2 P(h_i)]}{\sum_{i=1}^{n} -\log_2 P(h_i)}$$

$$\geq t \qquad \text{[from } \alpha_i \geq 0 \text{ and } P(h_i) < 1]$$

(ii) Suppose next $J(h_1, e), \ldots, J(h_n, e) < t$. Then, for $i = 1, \ldots, n$, there is some $\beta_i > 0$ such that:

$$J(h_i, e) = \frac{\log_2 P(h_i \mid e) - \log_2 P(h_i)}{-\log_2 P(h_i)}$$
$$= t - \beta_i$$

So,

$$\log_2 P(h_i|e) - \log_2 P(h_i) = (t - \beta_i)\,[-\log_2 P(h_i)] \qquad (3)$$

By plugging (3) into (1) above, we obtain:

$$
\begin{aligned}
J(h_1 \wedge \ldots \wedge h_n, e) &= \frac{\sum_{i=1}^{n}(t-\beta_i)[-\log_2 P(h_i)]}{\sum_{i=1}^{n} -\log_2 P(h_i)} \\
&= \frac{t\sum_{i=1}^{n} -\log_2 P(h_i) - \sum_{i=1}^{n}\beta_i[-\log_2 P(h_i)]}{\sum_{i=1}^{n} -\log_2 P(h_i)} \\
&= t - \frac{\sum_{i=1}^{n}\beta_i[-\log_2 P(h_i)]}{\sum_{i=1}^{n} -\log_2 P(h_i)} \\
&< t \qquad\qquad [\text{from } \beta_i > 0 \text{ and } P(h_i) < 1] \qquad \blacksquare
\end{aligned}
$$

## 4. Ordinal Equivalence

Suppose $J_1(h, e) = F_1(P(h|e), P(h))$ and $J_2(h, e) = F_2(P(h|e), P(h))$ are both continuous functions that are measures of justification. Then, they are ordinally equivalent to each other, i.e. for any two pairs $<h_i, e_i>$ and $<h_j, e_j>$, $J_1(h_i, e_i) < / = / > J_1(h_j, e_j)$ if and only if $J_1(h_i, e_i) < / = / > J_1(h_j, e_j)$.

**Proof:**

Let $\log_{P(h_i)} P(h_j) = r$, so that $[P(h_i)]^r = P(h_j)$. $r > 0$ because $0 < P(h_i), P(h_j) < 1$. Since $J_1(h, e) = F_1(P(h|e), P(h))$ and $J_2(h, e) = F_2(P(h|e), P(h))$ are continuous functions, it suffices to show that the claim holds for any two pairs $<h_i, e_i>$ and $<h_j, e_j>$ such that $[P(h_i)]^q = P(h_j)$ where $q$ is a positive rational number. Let $<m, n>$ be the smallest pair of positive integers such that $n/m = q$, so that $[P(h_i)]^n = [P(h_j)]^m$. Choose probabilistically independent (both unconditionally and conditionally on $e_i$*) propositions $h_1, \ldots, h_n$, and probabilistically independent (both unconditionally and conditionally on $e_j$*) propositions $h_{n+1}, \ldots, h_{n+m}$ such that:[16]

(i) $[P(h_i)]^n = [P(h_j)]^m$
(ii) $P(h_i) = P(h_1) = \ldots = P(h_n)$
(iii) $P(h_j) = P(h_{n+1}) = \ldots = P(h_{n+m})$
(iv) $P(h_i|e_i) = P(h_1|e_i^*) = \ldots = P(h_n|e_i^*)$
(v) $P(h_j|e_j) = P(h_{n+1}|e_j^*) = \ldots = P(h_{n+m}|e_j^*)$

[16] For example, think of $n$ urns of colored marbles, for each of which the probability of drawing a red marble is the same as $P(h_i)$, but given the evidence $e_i$* that the $n$ urns belong to a certain type, the probability of drawing a red marble is the same as $P(h_i|e_i)$. Similarly, think of $m$ urns of colored marbles, for each of which the probability of drawing a red marble is the same as $P(h_j)$, but given the evidence $e_j$*that the $m$ urns belong to a certain other type, the probability of drawing a red marble is the same as $P(h_j|e_j)$.

It follows from (ii) and (iv) that $J_1(h_i, e_i) = J_1(h_1, e_i^*) = \ldots = J_1(h_n, e_i^*)$. So, by SCR:

$$J_1(h_i, e_i) = J_1(h_1 \wedge \ldots \wedge h_n, e_i^*) \tag{1}$$

Similarly, it follows from (iii), (v) and SCR that:

$$J_1(h_j, e_j) = J_1(h_{n+1} \wedge \ldots \wedge h_{n+m}, e_j^*) \tag{2}$$

Since $h_1, \ldots, h_n$ are probabilistically independent, it follows from (ii) that $P(h_1 \wedge \ldots \wedge h_n) = [P(h_i)]^n$. Similarly, since $h_{n+1}, \ldots, h_{n+m}$ are probabilistically independent, it follows from (iii) that $P(h_{n+1} \wedge \ldots \wedge h_{n+m}) = [P(h_j)]^m$. But $[P(h_i)]^n = [P(h_j)]^m$ from (i). So,

$$P(h_1 \wedge \ldots \wedge h_n) = P(h_{n+1} \wedge \ldots \wedge h_{n+m}) \tag{3}$$

Meanwhile, since $h_1, \ldots, h_n$ are probabilistically independent conditionally on $e_i^*$, it follows from (iv) that $P(h_1 \wedge \ldots \wedge h_n | e_i^*) = [P(h_i|e_i)]^n$. Similarly, since $h_{n+1}, \ldots, h_{n+m}$ are probabilistically independent conditionally on $e_j^*$, it follows from (v) that $P(h_{n+1} \wedge \ldots \wedge h_{n+m} | e_j^*) = [P(h_j|e_j)]^m$. So,

$$\begin{aligned} &P(h_1 \wedge \ldots \wedge h_n|e_i^*) < / = / > P(h_{n+1} \wedge \ldots \wedge h_{n+m}|e_j^*) \\ &\qquad \text{iff } [P(h_i|e_i)]^n < / = / > [P(h_j|e_j)]^m \end{aligned} \tag{4}$$

Since $J_1(h, e) = F_1(P(h|e), P(h))$ is an increasing function of $P(h|e)$, it follows from (3) and (4) that:

$$\begin{aligned} &J_1(h_1 \wedge \ldots \wedge h_n, e_i^*) < / = / > J_1(h_{n+1} \wedge \ldots \wedge h_{n+m}, e_j^*) \\ &\qquad \text{iff } [P(h_i|e_i)]^n < / = / > [P(h_j|e_j)]^m \end{aligned} \tag{5}$$

It follows from (1), (2) and (5) that:

$$J_1(h_i, e_i) < / = / > J_1(h_j, e_j) \text{ iff } [P(h_i|e_i)]^n < / = / > [P(h_j|e_j)]^m \tag{6}$$

By the same reasoning,

$$J_2(h_i, e_i) < / = / > J_2(h_j, e_j) \text{ iff } [P(h_i|e_i)]^n < / = / > [P(h_j|e_j)]^m \tag{7}$$

It follows from (6) and (7) that:

$$J_1(h_i, e_i) < / = / > J_1(h_j, e_j) \text{ iff } J_2(h_i, e_i) < / = / > J_2(h_j, e_j) \qquad \blacksquare$$

## 5. Conjunction Theorem

Suppose $P(h_i)$, $P(h_j|h_i) \neq 1$. Then, $J(h_i \wedge h_j, e) > J(h_i, e)$ iff $J(h_j, e|\, h_i) > J(h_i, e)$.[17]

**Lemma 1:** $I(x \wedge y, z) = I(x, z|y) + I(y, z)$

**Lemma 2:** $I(x \wedge y) = I(x|y) + I(y)$.

**Corollary:** If (i) $P(h_1|e) < P(h_1)$ and (ii) $P(h_2|e \wedge h_1) > P(h_2|h_1)$, then $J(h_1 \wedge h_2|e) > J(h_1|e)$.

**Proof of Lemma 1:**

$$\begin{aligned}
I(x \wedge y, z) &= \log_2 P(x \wedge y|z) - \log_2 P(x \wedge y) \\
&= \log_2 P(x|y \wedge z)\, P(y|z) - \log_2 P(x|y)\, P(y) \\
&= [\log_2 P(x|y \wedge z) + \log_2 P(y|z)] - [\log_2 P(x|y) + \log_2 P(y)] \\
&= [\log_2 P(x|y \wedge z) - \log_2 P(x|y)] + [\log_2 P(y|z) - \log_2 P(y)] \\
&= I(x, z|y) + I(y, z) \qquad \blacksquare
\end{aligned}$$

**Proof of Lemma 2:**

$$\begin{aligned}
I(x \wedge y) &= -\log_2 P(x \wedge y) \\
&= -\log_2 P(x|y)\, P(y) \\
&= -[\log_2 P(x|y) + \log_2 P(y)] \\
&= I(x|y) + I(y) \qquad \blacksquare
\end{aligned}$$

**Proof of the Conjunction Theorem:**

$$\begin{aligned}
J(h_i \wedge h_j, e) - J(h_i, e) &= \frac{I(h_i \wedge h_j, e)}{I(h_i \wedge h_j)} - \frac{I(h_i, e)}{I(h_i)} \\
&= \frac{I(h_j, e \mid h_i) + I(h_i, e)}{I(h_i \mid h_j) + I(h_i)} - \frac{I(h_i, e)}{I(h_i)} \quad \text{[from Lemmas 1 and 2]} \\
&= \frac{[I(h_j, e \mid h_i) + I(h_i, e)]I(h_i) - I(h_i, e)[I(h_j \mid h_i) + I(h_i)]}{[I(h_j \mid h_i) + I(h_i)]I(h_i)} \\
&= \frac{I(h_j, e \mid h_i)I(h_i) - I(h_i, e)I(h_j \mid h_i)}{[I(h_j \mid h_i) + I(h_i)]I(h_i)}
\end{aligned}$$

But $I(h_j|h_i)$, $I(h_i) > 0$ from the assumption $P(h_i)$, $P(h_j|h_i) \neq 1$. So,

$$J(h_i \wedge h_j|e) > J(h_1|e) \quad \text{iff} \quad I(h_j, e|h_i)\, I(h_i) > I(h_1, e)\, I(h_j|h_i)$$

[17] $J(x, y|z) =_{\text{def}} I(x, y|z) / I(x|z)$ is the degree of justification for the proposition $x$ given the evidence $y$ on the background $z$. $I(x, y|z) =_{\text{def}} \log_2 P(x|y \wedge z) - \log_2 P(x|z)$ is the mutual information of $x$ given $y$ on the background $z$. $I(x|z) =_{\text{def}} -\log_2 P(x|z)$ is the self information of $x$ on the background $z$.

$$\text{iff} \quad \frac{I(h_j, e \mid h_i)}{I(h_j \mid h_i)} > \frac{I(h_i, e)}{I(h_i)}$$

$$\text{iff} \quad J(h_j, e|h_i) > J(h_i, e) \qquad \blacksquare$$

**Proof of the Corollary:**

$I(h_1, e) = \log_2 P(h_1|e) - \log_2 P(h_1) < 0$ from (i), while $I(h_1) > 0$ from the assumption. So,

$$J(h_1, e) = \frac{I(h_1, e)}{I(h_1)} < 0 \tag{1}$$

$I(h_2, e|h_1) = \log_2 P(h_2|e \wedge h_1) - \log_2 P(h_2|h_1) > 0$ from (ii), while $I(h_2|h_1) > 0$ from the assumption. So,

$$J(h_1, e \mid h_1) = \frac{I(h_1, e \mid h_1)}{I(h_2 \mid h_1)} > 0 \tag{2}$$

It follows from (1) and (2) that:

$$J(h_2, e|h_1) > J(h_1, e) \tag{3}$$

It follows from (3) by the Conjunction Theorem that $J(h_1 \wedge h_2|e) > J(h_1, e)$. $\blacksquare$

## References


Atkinson, D., 2009. "Confirmation and Justification," *Synthese*, this number.

Crupi, V., Fitelson, B., and Tentori, K. 2008. "Probability, Confirmation, and the Conjunction Fallacy," *Thinking and Reasoning* 14, pp. 182-199.

Crupi, V., Tentori, K., and Gonzalez, M. 2007. "On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues," *Philosophy of Science* 74, pp. 229-252.

Fitelson, B. 1999. "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity," *Philosophy of Science* 66, pp. S362-S378.

Fitelson, B. 2001. *Studies in Bayesian Confirmation Theory*. Ph.D. thesis, University of Wisconsin-Madison.

Gigerenzer, G. 1991. "How to Make Cognitive Ilusions Disappear: Beyond Heuristics and Biases," *European Review of Social Psychology* 2, pp. 83-115.

Hertwig, R. and Gigerenzer, G. 1999. "The 'Conjunction Fallacy' Revisited: How Intelligent Inferences Look Like Reasoning Errors," *Journal of Behavioral Decision Making* 12, pp. 275-305.

Huber, F. 2008a. "Assessing Theories, Bayes Style," *Synthese* 161, pp. 89-118.

Huber, F. 2008b. “Hempel’s Logic of Confirmation,” *Philosophical Studies* 139, pp. 181-189.

Kyburg, H. 1961. *Probability and the Logic of Rational Belief* (Middletown: Wesleyan University Press)

Makinson, D. C. 1965. “The Paradox of the Preface,” *Analysis* 25, pp. 205-207.

Marr, D. 1982. *Vision* (New York: W. H. Freeman and Company).

Meijs, W. 2005. *Probabilistic Measures of Coherence*. Ph. D. dissertation, Erasmus University, Rotterdam.

Reichenbach, H. 1956. *The Direction of Time* (Berkeley: University of California Press).

Shafir, E., Smith, E., and Osherson, D. “Typicality and Reasoning Fallacy,” *Memory & Cognition* 18, pp. 229-239.

Shogenji, T. 1999. “Is Coherence Truth Conducive?” *Analysis* 59, pp. 338-345.

Sides, A. Osherson, D., Bonini, N., and Viale, R. 2002. “On the Reality of the Conjunction Fallacy,” *Memory & Cognition* 30, pp. 191-198.

Tversky, A. and Kahneman, D. 1983. “Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment,” *Psychological Review* 90: 4, pp. 293-315.

Williamson, T. 2000. *Knowledge and Its Limits* (Oxford: Oxford University Press).