Stalnaker's Hypothesis: A Critical Examination of Hájek's Counter Argument*

Mehmet Hilmi Demir

Abstract: According to what is known as Stalnaker's hypothesis, the probability of a conditional statement is equal to the conditional probability of the statement's consequent given the statement's antecedent. Starting with David Lewis, many have attempted to show that this hypothesis cannot be true for non-trivial probability functions. These attempts, which are known as the triviality results, cannot refute the hypothesis conclusively, because the triviality results usually rest on controversial assumptions such as the closure of conditionalization. In addition to the triviality results, there is one often cited argument against Stalnaker's hypothesis that does not seem to rest on a controversial assumption. The argument is Alan Hájek's 1989 reductio argument, which purportedly shows that Stalnaker's hypothesis leads to outright contradiction. In this paper, I critically evaluate Hajek's reductio argument and show that it is not a valid argument. His argument is simply an instance of the *petitio principii* fallacy. On the positive side, I argue that my critical evaluation of Hajek's argument brings us one step closer to the reconciliation of the analytical and empirical examinations of Stalnaker's hypothesis.
Keywords: Conditional Probability; Probability of a Conditional Statement; Triviality results; Stalnaker; Hájek.

Özet: Literatürde Stalnaker hipotezi olarak bilinen iddiaya göre, bir şartlı önermenin olasılığı, o önermenin art bileşenin ön bileşeninine şartlı olasılığına eşittir. David Lewis'in 1976 tarihli makalesinden beri birçok felsefeci bu iddianın sadece basit ve sıradan (trivial) olasılık fonksiyonları için geçerli olduğu, diğer daha işlevli (non-trivial) olasılık fonksiyonlarına uygulanamayacağını göstermeye çalışmışlar ve bu hedef doğrultusunda birçok ispat sunmuşlardır. Ancak sıradanlık sonuçları (triviality results) olarak bilinen bu tür ispatların Stalnaker hipotezini tam olarak reddetmeye yeterli olmadığı anlaşılmıştır. Çünkü bu ispatların büyük bir çoğunluğu koşullamanın kapalılığı (closure of conditionalization) gibi tartışmalı olan varsayımlara dayanmaktadır. Literatürde tartışmalı herhangi bir varsayıma dayalı olmadığı iddia edilen ve sıklıkla gönderme yapılan bir başka argüman daha mevcuttur. Alan Hájek'in 1989 tarihli makalesinde olmayana ergi metodu ile geliştirdiği bu argüman, herhangi tartışmalı bir varsayıma dayanmadan,

Demir, Mehmet Hilmi. (2017). Stalnaker's Hypothesis: A Critical Examination of Hájek's Counter Argument. *Kilikya Felsefe Dergisi* (2). pp. 30-40.

*Makalede yer alan kimi semboller derginin matbu versiyonunda dizgiden kaynaklı hatalı basılmıştır.

Stalnaker hipotezinin doğrudan çelişkiye neden olduğunu göstermektedir. Bu makalede Hájek'in argümanının geçerliliği detaylı olarak incelenmekte ve sonuçta söz konusu argümanın *petitio principii* çıkarsama hatasını barındırdığı ve bu sebeple de geçerli olmadığı tespit edilmektedir. Pozitif katkı olarak ise bu varılan tespitin Stalnaker hipotezinin analitik ve ampirik değerlendirmeleri arasında var olan uyuşmazlığın giderilmesinde bir adım daha ileri gitmemizi sağladığı iddia edilmektidir.

Anahtar kelimeler: Şartlı Olasılık; Şartlı Önermelerin Olasılığı; Sıradanlık Sonuçları; Stalnaker; Hájek.

Introduction

According to what is known as Stalnaker's hypothesis[1], the probability of a conditional statement is equal to the conditional probability of the statement's consequent given the statement's antecedent. That is,

$$Pr\ (A \rightarrow C) = Pr\ (C \mid A) \text{ whenever } Pr\ (A) > 0.$$

This hypothesis, which was offered in the 1960s, has been criticized extensively. Starting with the paper by Lewis (1976) that is now a classic, "Probabilities of Conditionals and Conditional Probabilities", many have offered what is known as the triviality results against this hypothesis[2]. More or less, the triviality results show that Stalnaker's hypothesis (henceforth **SH**) holds only for very trivial probability functions; it fails for all the other more interesting and more relevant probability functions. It was, however, observed that all of the triviality results rest on at least one of the following two assumptions (Stalnaker, 1976; Hájek, 1989):

A1. Each class of probability functions is closed under certain operations (such as conditionalization or Jeffrey conditionalization).

A2. The proposition expressed by a conditional sentence is independent of the probability function defined on it.

[1] After Lewis' triviality results, Stalnaker lost faith in his hypothesis. In a letter dated January 1974 to van Fraassen, he says the following: "In fact, I am as taken with the distinction between the probability of the conditional and the conditional probability as I once was with their supposed identity" (available in Van Fraassen 1976, p.305). Despite this change in Stalnaker's position, the hypothesis is still called Stalnaker's hypothesis in the literature. In this paper, I follow this tradition and refer to the hypothesis as Stalnaker's hypothesis.

[2] For some examples of triviality results, see Lewis (1976, 1986), Hajek (1989, 1994, 2011), and Döring (1994).

If these two assumptions are not true in general, then it is fair to conclude that **SH** is immune to the damaging triviality results[3]. A defense of **SH** along these lines is provided by Van Fraassen. In formal terms, Van Fraassen (1976, 1989) showed that there is no inconsistency in accepting **SH** while rejecting **A1** and/or **A2**. The possibility of such a defense, as recently observed by Dietz and Douven (2011), implies that the triviality results cannot conclusively refute **SH**.

There are, however, other arguments that purportedly refute **SH** and do not rely on any of the assumptions stated above. Perhaps the most famous among those arguments[4] is the one offered by Hájek in his 1989 article, "Probabilities of Conditionals Revisited". In his paper, Hájek claims to prove that **SH** leads to outright contradiction. The only assumption that Hájek requires for his proof is that ($A \rightarrow C$) is a proposition whenever A and C are. Since it is quite difficult to think of a less controversial assumption than this, **SH** cannot be defended against Hájek's proof by questioning the underlying assumption. Do we have to conclude that **SH** is conclusively refuted by Hájek's reductio proof? In this paper, I examine Hájek's proof in order to answer this question. If Hájek's proof is valid and non-circular, then the answer will be yes. But as I show below, Hájek's proof is at best a circular one. One of the steps that Hájek uses in his reductio proof amounts to the negation of **SH**. With this step, he then proceeds to prove that **SH** is false. This makes Hájek's proof circular. Given this problematic nature of Hájek's proof and the availability of Van Fraassen's line of defense against the triviality results, I conclude that, contrary to what is commonly accepted in the literature, the Stalnaker hypothesis has not been conclusively refuted.

The conclusion of this paper not only defends Stalnaker's hypothesis against Hájek's argument, but also serves the purpose of resolving a puzzling tension that exists between the analytical results and the empirical findings about **SH**. It has empirically been shown that human judgments of the probability of conditional statements fit quite well with

[3] Hajek (2011) recently offered another triviality result against **SH.** This most recent triviality result also rests on one of the aforementioned controversial assumptions, i.e. **A1**.

[4] Douven and Dietz (2011) also have an argument against **SH** that does not rest on **A1** or **A2**. In their argument, they rightly point out that **SH** implies the probabilistic independence of the antecedent of a conditional statement from the conditional statement itself. They find this implication "deeply problematic" and "plain wrong". Their argument that aims to show the implausibility of the probabilistic independence thesis, however, is not a valid one. The examination of Douven and Dietz's argument is beyond the scope of this paper.

the predictions of **SH**[5]. In other words, humans judge the probability of a conditional statement to be equal to the probability of the conditional statement's consequent given its antecedent. But, as briefly mentioned above, because of the triviality results and arguments such as Hájek's, it is generally accepted that analytical results refute **SH**. This means that there is a puzzling tension between the empirical and analytical analyses of **SH**. The conclusion of this paper helps to remove one of the arguments that lead to this puzzling tension.

The paper starts with a reconstruction of Hájek's proof. The circular nature of Hájek's proof is then shown. The paper is concluded with a positive discussion about the consistency between the empirical and analytical analyses of Stalnaker's hypothesis.

2. A Step-by-Step Reconstruction of Hjek's Proof

Hájek's elegant proof rests on quite a simple idea. Since **SH** equates *Pr (C|A)* to *Pr (A→C)*, for every conditional probability there must be a matching unconditional probability simply because *Pr (A→C)* itself is an unconditional probability. Thus, finding a conditional probability that does not have any matching unconditional probability will be sufficient enough for showing that **SH** is false. This is the strategy that Hájek uses in his proof. His proof, which applies only to finite models with at least three possible worlds[6], assumes that there is an unconditional probability for every conditional probability, and then attempts to derive a contradiction. Under this reductio assumption, Hájek, without loss of generality, assumes that the probabilities of possible worlds are non-zero and arranges those probabilities in an increasing order. After that, his proof proceeds with two different cases. In *Case 1*, the highest value in the ordered probability series is less than ½, and in *Case 2*, the highest value is greater than or equal to ½. Since the proof in both cases has the same structure with little change in arithmetical details, I examine only *Case 1*. The following is a step-by-step reconstruction of Hájek's proof for *Case 1*.

Step 0: *Every conditional probability equals some unconditional probability.*

[5] Some examples of the empirical investigation of **SH** are the works of Hadjichristidis et al. (2001), Evans et al. (2003), Oberauer and Wilhelm (2003), Over and Evans (2003), Evans and Over (2004), and Over et al. (2007).

[6] Hájek's proof does not apply to models with one or two possible worlds because of the simplicity of such models. In such models, **SH** is trivially true.

This is the reductio assumption.

Step 1: $0 < p_1 \leq p_2 \leq p_3 \leq \cdots \leq p_n$ *where* $\sum_i^n p_i = 1$, *and* $p_n < \frac{1}{2}$

Without loss of generality, Hájek assumes that there are *n* possible worlds and the probability of each of those worlds is non-zero. It is also possible to put those probabilities, whose sum is *1*, in an increasing order. The second part of this step states that the highest value in the probability series is less than ½. This is the *Case 1* that Hájek considers in his proof.

$$\textbf{\textit{Step 2}}: \; p_1 \leq \frac{p_1}{1-p_2} \leq \frac{p_1}{1-p_3} \leq \cdots \leq \frac{p_1}{1-p_n} \leq 2p_1 \leq p_1 + p_2$$

This is the case because $p_i \leq p_{i+1}$, as stated in Step 1. There are *n-1* values of the form $\frac{p_1}{1-p_i}$ because in the ordered series above *i* is between *2* and *n*, inclusive.

Step 3: *These n-1 values of the probabilities of the form* $\frac{p_1}{1-p_i}$ *are conditional probabilities.*

The probabilities of this form are indeed conditional probabilities. To give an example, let's assume that W_1, W_2, and W_3 are the only possible worlds with the probability values of p_1, p_2, and p_3. Let A, B, and C be the only statements that can be evaluated in these possible worlds. Further assume that only C is true in W_1; all three statements are true in W_2; and only A is true in W_3. In this toy example, the value $\frac{p_1}{1-p_2}$ is equal to the conditional probability of C given not-B.

Step 4: *These values need to be matched with unconditional probabilities strictly between* p_1 *and* p_1+p_2.

This follows from Steps 2 and 3.

Step 5: *The only unconditional probabilities between* p_1 *and* p_1+p_2 *are* $p_2, p_3, \ldots, p_n$.

This step is the most crucial one in Hájek's proof. At first glance, it may seem that this straightforwardly follows from Step 1, where the probabilities of possible worlds are put in an increasing order. However, it should be noted that Step 1 only states the set of the probabilities of possible worlds, and this set does not necessarily limit the set of all unconditional probabilities. Thus, this step requires further justification, and unfortunately, Hájek does not provide any.

Step 6: $\frac{p_1}{1-p_2} = p_2, \quad \frac{p_1}{1-p_3} = p_3, \quad \dots \quad \frac{p_1}{1-p_n} = p_n$

This follows from Steps 2 and 5.

Step 7: $p_2 = p_3 = \cdots = p_n$

This follows from solving the equation $\frac{p_1}{1-p_i}$ $=p_i$, which is established in Step 6.

Step 8: *There can be no unconditional probability strictly between p_1 and p_1+p_2.*

This is a direct consequence of the previous step.

Step 9: *There is a conditional probability between p_1 and $p+p_2$, namely* $\frac{p_2}{1-p_1}$.

$\frac{p_2}{1-p_1}$ is indeed a conditional probability. Just consider the conditional probability of (B and C) given A in the toy example given in Step 3.

Step 10: *Contradiction.*

Steps 8 and 9 obviously contradict with each other.

Thus: *Reductio assumption is false.*

Corollary:

Since **SH** implies the reductio assumption and the reductio assumption is false, **SH** does not hold for any finite model of probability with at least 3 worlds.

3. The Form of Hájek's Proof

Let me use the following symbols for the ease of presentation: **RA** for the reductio assumption in Step 0 and **X** for the crucial premise in Step 5. It is worth noting that the premise in Step 5 is independent of **RA**. In other words, **RA** does not imply **X**. All other steps in Hájek's proof follow from previous steps by simple algebraic manipulations. Thus, the form of the proof is as follows.

RA

....

....

X

....

....

Contradiction.

Therefore not-RA.

If **X** were an implication of **RA**, then this proof form would be valid. But it is not. Knowing that every conditional probability has a corresponding unconditional probability does not tell us anything about the set of unconditional probabilities between p_1 and $p_1 + p_2$. The other way of making this proof form valid is to give independent justification for the truth of **X**. No such explanation is provided. Moreover, a careful examination shows that accepting **X** amounts to rejecting **SH**. **X** states that the only unconditional probabilities that could be strictly between p_1 and $p_1 + p_2$ are $p_2, p_3, p_4, \ldots, p_n$. If **SH** is true, however, then there are many more unconditional probabilities in that interval, because every value of conditional probabilities will also be the value of an unconditional probability[7]. For example, if **SH** is true, the value $\frac{p_2}{1-p_1}$, which is a conditional probability, is also an unconditional probability, because $Pr(A \rightarrow C)$ itself is an unconditional probability.

[7] At this point, in defense of Hájek's argument, one may reason in the following way : The only unconditional probabilities that could be strictly between p_1 and $p_1 + p_2$ are $p_2, p_3, p_4, \ldots, p_n$, because unconditional probabilities are probabilities of propositions which are sets of worlds. And the probability of a proposition is the sum of the probabilities of the worlds that constitute it. So, all propositions must have probabilities that are sums of the p_i. But, if we assume that **SH** is true then the probability of a conditional statement, which is itself a proposition, will not necessarily be a probability of a set of worlds. Thus, to claim that it is generally accepted that conditionals correspond to sets of worlds and this justifies Hájek's argument will simply be stating the negation of **SH**, and thus, will again be an instance of *petitio principii*. I thank Alan Hájek for raising this issue (personal communication, September 30, 2012).

The only remaining way of making Hájek's proof valid is to change its conclusion. The conclusion that Hájek could draw from his proof is not-(**RA** and **X**). But if this is the correct conclusion of the proof, then the corollary about **SH** does not follow, simply because of the fact that the following set of propositions is consistent: **SH → RA**, not-(**RA** and **X**), **SH**. Just assign True to **SH** and **RA**, and False to **X**.

4. Discussion and Concluding Remarks

The empirical investigation of how humans judge the probability of a conditional statement more or less started in the late 1990s. Since then, various experimental psychologists have conducted experiments for understanding the human judgment of the probability of conditional statements. Almost all of those experiments have found that the best predictor of the judgment of the probability of a conditional statement is the conditional probability of the conditional statement's consequent given its antecedent. This is nothing but what Stalnaker's hypothesis asserts, i.e. $Pr\ (A \rightarrow C) = Pr\ (C|A)$ whenever $Pr\ (A)$ is not zero. In other words, **SH** has been subjected to extensive empirical testing. In order to emphasize the thoroughness of those empirical tests, it is worth briefly mentioning a set of experiments that were conducted for testing **SH**.

The early empirical tests of **SH** mostly used what is called basic conditionals in the experimental procedures. Basic conditionals are conditionals that have a neutral content; that is to say, their contents are as independent as possible from context and background knowledge. "If the card is yellow, then it has a circle on it" is such a conditional. However, conditionals that are used for decision making and planning are rarely basic conditionals. Thus, the earlier experiments account only for a limited and idealized subset of conditionals. Over et al. (2007), in their three experiments, used non-basic indicative conditionals and related counterfactuals. An example of a non-basic conditional that they used is: if the cost of petrol increases, then traffic congestion will improve. In that respect, their findings apply to a larger domain. In their experiments, they tested four different hypotheses about the probability judgments of a conditional statement. The first of those is the conjunctive probability hypothesis, which equates the probability of a conditional statement to the probability of its antecedent and the consequent. The second hypothesis is the material conditional hypothesis, which treats the probability of a conditional statement, say if A then B, as the probability of the truth functional equivalent of the conditional statement, which is (not-A or B). The third is what Over et al. call the conditional probability thesis, which is identical to **SH**. The last one is the hypothesis that the probability of a conditional statement is proportional to the extent to which the antecedent's probability raises the probability of the consequent, which is known as *the delta-p rule*. Each one of these four hypotheses is an implication of a different theory of conditionals. Moreover, these four hypotheses span the entire space of alternative hypotheses available in the literature. Over et al. test these hypotheses under different conditions and their findings show that the conditional probability hypothesis, i.e. Stalnaker's hypothesis, is by far the best predictor of the experimental subjects' judgments of the probability of conditional statements. Their findings imply that the conditional hypothesis is empirically verified not only for indicative conditionals but also for counterfactuals.

Robust findings such as those of Over et al. provide strong support for **SH**, but as we have seen above, there are analytical arguments that purportedly show that the hypothesis cannot be true. This situation presents quite a puzzling tension. In the face of this tension, Douven and Dietz rightly issue a call for help in their 2011 paper. What should we do? It seems that one could either downplay the significance of the empirical findings by trying to explain them away or one could take those findings seriously and go back to the drawing board. The latter option seems to be the better one in its potential to leading to new analytical findings or revising old ones.

The puzzling situation that we have here has a strong resemblance to the tension that was pointed out by Tversky and Kahnemann (1983) for the way humans judge the probability of conjunctive statements. As is well known by now, the empirical investigation of probability judgments showed us that humans, at least in some cases, assign a higher value to the probability of a conjunction of two statements than the probability value assigned to one of the conjuncts; this is the so-called conjunction fallacy. This finding contradicts with the classical Kolmogorovian probability axioms, which imply that the probability of a conjunction is less than or equal to the probability of one of the conjuncts alone. Taking such empirical findings seriously instead of downplaying them turned out to be the better option, because it led to new theories in modeling human probability judgment. To give an example, in a recent book, Busemeyer and Bruza (2012, p. 16-18) convincingly argue that quantum probability axioms provide a better model (at least in some cases) in explaining human probability judgments of conjunctions, because in quantum probability the probability of A and B under certain circumstances can be higher than the probability of A.

In a similar spirit, we should take the empirical findings about **SH** seriously and go back to the drawing board. At the drawing board, the first thing that needs to be done is to verify the validity of the analytical arguments offered against **SH**. This is what I have set out to do in this paper. My scrutiny of Hájek's reductio argument, which does not seem to rest on controversial assumptions, shows that his argument is at best a circular one. In other words, his argument is at best an instance of the fallacy of *petitio principii*. With Hájek's reductio argument out of the way, it is fair to conclude that we are one step closer to the reconciliation of the empirical and analytical findings about Stalnaker's hypothesis. Needless to say, there is still a long way to go.

Mehmet Hilmi Demir, Orta Doğu Teknik Üniversitesi, Türkiye

References

Busemeyer, J.R., & Bruza, P.D (2012). *Quantum Models of Cognition and Decision*. Cambridge University Press: Cambridge, UK.

Douven, I. & Dietz, R. (2011). A puzzle About Stalnaker's hypothesis. *Topoi*, 30, 31-37.

Döring, F. (1994). On the probabilities of conditionals. *Philosophical Review,* 103, 689-699.

Evans, J.St.B.T., Handley, S.J., & Over, D.E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 321-355.

Evans, J.St.B.T., & Over, D.E. (2004). *If*. Oxford University Press: Oxford, UK.

Hájek, A. (1989). Probabilities of conditionals—revisited. *Journal of Philosophical Logic,* 18, 423-428.

Hájek, A. (1994). Triviality on the cheap?. In: Eells, E., Skyrms, B. (eds.) *Probability and Conditionals*. Cambridge University Press: Cambridge, UK, pp. 113-140.

Hájek, A. (2011). Triviality pursuit. *Topoi*, 30, 3-15.

Hadjichristidis, C., Stevenson, R.J., Over, D.E., Sloman, S.A., Evans, J.St.B.T., & Feeney, A. (2001) On the evaluation of 'if p then q' conditionals. In: Moore, J.D., Stenning, K. (eds.) *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*. Edinburgh, pp. 381-386.

Lewis, D.K. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85, 297-315.

Lewis, D.K. (1986). Probabilities of conditionals and conditional probabilities II. *Philosophical Review,* 95, 581-589.

Oberauer, K., & Wilhelm, O. (2003). The meaning(s) of conditionals: conditional probabilities, mental models and personal utilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29, 688-693.

Over, D.E., & Evans, J.St.B.T. (2003). The probability of conditionals: the psychological evidence. *Mind and Language,* 18, 340-358.

Over, D.E., Hadjichristidis, C., Evans, J.St.B.T., Handley, S.J., & Sloman, S.A. (2007). The probability of causal conditionals. *Cognitive Psychology*, 54, 62-97.

Stalnaker, R. (1968). A theory of conditionals. In: Rescher, N. (ed.) *Studies in Logical Theory* (*American Philosophical Quarterly* Monograph Series, No. 2). Blackwell, Oxford, pp. 98-112.

Tversky, A. & Koehler, D.J. (1983) "Extensional versus intuitive reasoning: the conjunctive fallacy in probability judgment". *Psychological Review*, 90, 293-315.

Van Fraassen, B. (1976) "Probabilities of conditionals". In Harper, W.L. and Hooker, C.A. (eds.) *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. I. Reidel: Dordrecht: pp. 261-308.

Van Fraassen, B. (1989) "Rationality does not require conditionalization". In Ullmann-Margalit, E. (ed.) *The Israel Colloquium: Studies in the History, Philosophy and Sociology of Science*, Vol. 5. Kluwer Academic Publishers: Dordrecht.