The Two-envelope Paradox With No Probability[*]

Byeong-Uk Yi

Department of Philosophy

University of Toronto

Toronto, ON M5R 2M8

b.yi@utoronto.ca

Smullyan (1992, 189-192) formulates an intriguing version of the two-envelope paradox. The usual versions of the paradox invoke decision-theoretic principles that relate expected values to rationality of choice. Smullyan's invokes no such principles. So one cannot resolve it by attending to illicit assumptions about probability or rationality of choice. In this paper, I present an analysis of, and solution to, Smullyan's version of the paradox by attending to the logic of conditionals. The analysis and solution yield and help to highlight interesting results about the logic of conditionals.

1. Smullyan's Paradox

Consider a usual formulation of the two-envelop paradox:

> There are two envelopes, Ali and Baba, on a table, and you know that one of them has twice

[*]I wish to thank Adam Murray and anonymous referees of *The Journal of Philosophy* for useful comments on earlier versions of this paper.

> as much money as the other, but not which one has more. Is it rational to choose one of them over the other? One might argue that it is as follows. Suppose that Ali has \$$x$. Then Baba is equally likely to have \$2x as to have \$$x$/2. So its expected value is \$1¼x (=½ x \$2x + ½ x \$$x$/2), which is greater than the amount in Ali, \$$x$. So Baba is preferable.

This yields a paradox or at least a puzzle: the reasoning might seem to be correct, but the same reasoning leads to the opposite conclusion, that Ali is preferable, as well. The reasoning invokes some assumptions about probability distribution and decision theoretic principles that relate expected values to rationality of choice. So one might attempt to resolve the paradox by rejecting wrong assumptions about probability or rationality of choice.[1] Smullyan formulates a related paradox that invokes no such assumptions.

To do so, he considers the following situation (call it *the Smullyan situation*):[2]

> There are two sealed envelopes on the table. You are told one of them contains twice as much money as the other. . . . You pick up one of the two envelopes and decide that you are going to trade it for the other. (*Ibid*., 189)

---

[1]For more on the usual, probabilistic versions of the two-envelope paradox, see, e.g., Nalebuff (1988) & (1999), Jackson et al. (1994), Bloome (1995), Chalmers (2002), Priest & Restall (2008), and Yi (2009). In Yi (2009), I argue that a proper solution of the paradox requires a radical departure from the standard, probability-based decision theories.

[2]Note that the descriptions of the situation do not include the condition that the amounts in the envelopes are not known. It is not necessary to invoke this condition in formulating Smullyan's version of the two envelope paradox. See Smullyan's arguments for Propositions *1* & *2* given below.

And he presents arguments for two contradictory theses on the situation: “Proposition *1*. The amount that you will gain, if you do gain, is greater than the amount you will lose, if you lose”, and “Proposition *2*. The amounts are the same” (*ibid*., p. 190). Here is his argument for Proposition *1*:

> Let n be the number of dollars in the envelope you are now holding. Then the other envelope has either 2n or n/2 dollars. . . . Then if you gain on the trade, you will gain n dollars, but if you lose on the trade, you will lose n/2 dollars. Since n is greater than n/2, then the amount you gain, if you do gain–which is n–is greater than the amount you will lose, if you do lose–which is n/2. (*Ibid*., 190f)

And his argument for Proposition *2*:

> Let d be the difference between the amounts in the two envelopes . . . . If you gain on the trade, you will gain d dollars, and if you lose on the trade, you will lose d dollars. And so the amounts are the same after all. (*Ibid*., 191)

Because the Smullyan situation is neither inconsistent nor impossible, it cannot satisfy both propositions.[3] So at least one of the arguments must be faulty. What is wrong with them?

Clearly, one cannot solve this version of the two-envelop paradox (call it *Smullyan’s paradox*) by examining probability distributions or disputing some probabilistic principles about

[3]I say that a situation (or possible world) *satisfies* a statement, if the statement holds on the situation. I use ‘hold *on*’ rather than ‘hold *in*’ to avoid suggesting that a statement satisfied by a situation (or possible world) must be *in* the situation (or possible world).

rationality of choice. Some might think it is necessary to give a proper account of decisions and their results to solve the paradox, because Smullyan's formulation of it involves talks of the *decision* to trade envelopes, and the *gain* or *loss* that might result from it. But we can formulate it with no reference to decisions or their results.

Imagine a situation, *S*, in which there are two envelopes, *Ali* and *Baba* (in short, *a* and *b*), on a table, and one of them has twice as much money as the other.[4] Let f(a) and $f(b)$ be the amounts in dollars in Ali and Baba, respectively. Then we can replace Propositions *1* & *2* with the following theses about the difference between the two amounts:

[P1] The difference between $f(a)$ and $f(b)$ in case Baba has more than Ali *is greater than* the difference between $f(a)$ and $f(b)$ in case Ali has more than Baba.

[P2] The difference between $f(a)$ and $f(b)$ in case Baba has more than Ali *is the same as* the difference between $f(a)$ and $f(b)$ in case Ali has more than Baba.

And we can turn Smullyan's arguments for those propositions to arguments for these theses:

Argument 1: Let $f(a)$ be a positive number n. Then $f(b)$ is either $2n$ or $n/2$. Then if Baba has more than Ali, the difference between $f(a)$ and $f(b)$ is n; but the difference is $n/2$, if Ali has more than Baba. So the difference between $f(a)$ and $f(b)$ if Baba has more (i.e., n) is greater than the difference between them if Ali has more (i.e., $n/2$).

[4]We may take this to imply that both have some money, i.e., a positive amount of money.

Argument 2: Let *d* be the difference between *f*(*a*) and *f*(*b*). If Baba has more than Ali, the difference is *d*; and the difference is *d*, again, if Ali has more than Baba. So the differences in the two cases are the same.

These arguments cannot both be correct, because *S* is clearly a possible situation. What is wrong with them?

2. Analysis of Smullyan's Paradox

To assess Arguments *1* & *2,* it is necessary to divide both of them into two parts. Their first parts lead to the following theses:

[T1] There is a number *n*, and a number *m* smaller than *n* such that if Baba has more than Ali, the difference between *f*(*a*) and *f*(*b*) is *n,* whereas if Ali has more than Baba, the difference is *m*.

[T2] There is a number *n* such that if Baba has more than Ali, the difference between *f*(*a*) and *f*(*b*) is *n*, and that if Ali has more than Baba, the difference is also *n*.

The second parts consist in inferring [P1] from [T1], and [P2] from [T2]. Both arguments, as we shall see, break down in the second parts. That is, it is wrong to infer [P1] and [P2] from [T1] and

[T2]. These are compatible[5] because they are both implied by the descriptions of situation *S*, but they, taken together, contradict both [P1] and [P2].[6]

To see this, it is useful to formulate the descriptions of the situation as well as the four theses, [T1]–[T2] & [P1]–[P2], in a regimented language. To formulate them in an elementary language that includes expressions for numbers and usual operations on them, let '*k*', '*m*', and '*n*' be restricted variables for numbers, and '*x*','*y*', and '*z*' restricted variables for the envelopes on the table; and let '*a*' and '*b*' be singular constants, and '*C*' a dyadic predicate whose meanings are given as follows:

a: Ali

b: Baba

xCn: *x* has exactly *n* dollars.

---

[5]I say two or more statements are *compatible* (on *w*), if there is a possible situation (accessible from *w*) that satisfies all of them. So ϕ and ψ are compatible if and only if $\diamond(\phi \,\&\, \psi)$ holds (on *w*).

[6]In his analysis of Smullyan's paradox, Chase (2002) formulates Proposition *1* as follows:

(C) There are *x* and *y* such that . . . (1) if you gain, you gain $*x* . . . (2) if you lose, you lose $*y*, and . . . (3) *x* > *y*. (*Ibid.*, 158)

And he argues that this cannot be satisfied by any possible situation (e.g., the Smullyan situation). For a proper analysis of the paradox, however, it is necessary to distinguish Proposition *1* from Chase's thesis (C), which amounts to [T1], not [P1]. The Smullyan situation satisfy (C) while violating Proposition *1*, as we shall see, as *S* satisfies [T1] while violating [P1]. See the last paragraph of this section. Incidentally, Chase holds that the conditionals in (C) must be considered counterfactual conditionals, but there is no reason to do so to formulate Smullyan's paradox although one might do so to formulate its counterfactual cousin. See the first paragraph of section 4.

And let '$f(x)$', '$d(x, y)$', and '$x \succ y$' abbreviate '$(\iota n)xCn$',[7] '$|f(x) - f(y)|$', and '$f(x) > f(y)$', respectively. Then '$f(x)$' and '$d(x, y)$' refer to the dollar amount in x, and the difference between the dollar amounts in x and y, respectively; and '$x \succ y$' means *x has more money than y*.

Now, we can formulate the conditions used to describe situation S as follows:[8]

(0) $\exists x \exists y[\forall z(z = x \vee z = y) \,\&\, f(x) = 2 \cdot f(y) \,\&\, f(y) > 0]$. (There are exactly two envelopes on the table, and one of them has twice as much money as the other.)

(1) $\exists x\, x=a \,\&\, \exists x\, x=b \,\&\, a \neq b$. (Ali is one of the envelopes on the table, Baba is also one of them, and Ali is not Baba.)

These conditions are consistent and compatible, and there are possible situations that satisfy both of them. (S is one of those situations.) And [T1] and [T2] can be formulated as follows:[9]

[T1] $\exists n \exists m < n([b \succ a \rightarrow d(a, b) = n] \,\&\, [a \succ b \rightarrow d(a, b) = m])$. (There is a number n, and a number m smaller than n such that if Baba has more than Ali, the difference between $f(a)$ and $f(b)$ is n, whereas if Ali has more than Baba, the difference is m.)

[7] I use the Greek letter iota 'ι' as the definite description operator.

[8] The two conditions, (0) and (1), in a sense boil down to just one, (0). Assuming (1) while using two proper names, 'Ali' (or 'a') and 'Baba' (or 'b'), facilitates formulation of Smullyan's paradox, but one can formulate it without using the names. One might argue that the existential generalization of the conjunction of [P1] and [P2], given below, follows from (0) because it follows from the existential generalization of the conjunction of [T1] and [T2], given below, while this follows from (0). (Note that the existential generalization of (1) follows from (0).)

[9] I use '$\rightarrow$' for the indicative conditional. I do not assume the usual analysis of the indicative conditional as the material conditional, for which I use '$\supset$'. See the discussion in section 3.

[T2] $\exists n([b \succ a \rightarrow d(a, b) = n] \;\&\; [a \succ b \rightarrow d(a, b) = n])$. (There is a number n such that if Baba has more than Ali, the difference between $f(a)$ and $f(b)$ is n, and that if Ali has more than Baba, the difference is also n.)

To formulate [P1] and [P2] in an elementary language, it is necessary to use the definite descriptions '$(\iota k)[b \succ a \rightarrow d(a, b) = k]$' and '$(\iota k)[a \succ b \rightarrow d(a, b) = k])$', which amount to 'the difference between $f(a)$ and $f(b)$ in case Baba has more than Ali' and 'the difference between $f(a)$ and $f(b)$ in case Ali has more than Baba', respectively. Using those descriptions, we can formulate the theses as follows:

[P1] $(\iota k)[b \succ a \rightarrow d(a, b) = k] > (\iota k)[a \succ b \rightarrow d(a, b) = k])$.

[P2] $(\iota k)[b \succ a \rightarrow d(a, b) = k] = (\iota k)[a \succ b \rightarrow d(a, b) = k])$.

We can now see that (0)–(1) imply both [T1] and [T2]. To see this, note that (0)–(1) imply the following:[10]

(2) $\exists x \exists y[\forall z(z = x \vee z = y) \;\&\; x \succ y]$. (There are exactly two envelopes on the table, and one of them has more money than the other.)

(3)a. $\exists n(n > 0 \;\&\; f(a) = n)$. (Ali has a positive amount.)

b. $\exists n(n > 0 \;\&\; f(b) = n)$. (Baba has a positive amount.)

(4) $\forall x \forall y[x \succ y \rightarrow f(x) = 2 \cdot f(y)]$. (If an envelope on the table has more than another, the amount in the former is twice that of the amount in the latter.)

[10] Note that (1)–(4), which imply (0), are equivalent to (0)–(1).

Now, (1)–(4) imply [T1] and [T2]. [T1] follows from the following:

[T1a] $\exists n > 0([b \succ a \rightarrow d(a, b) = n] \,\&\, [a \succ b \rightarrow d(a, b) = n/2])$.

And this follows from (1), (3a), and (4):[11]

*Argument 1**: Let a number n greater than 0 be $f(a)$ (there is such a number by (3a)). Then if $b \succ a$, then $f(b) = 2n$ (by (1) & (4)) and $d(a, b) = f(b) - f(a) = 2n - n = n$. And if $a \succ b$, then $f(b) = n/2$ (by (1) & (4)) and $d(a, b) = f(a) - f(b) = n - n/2 = n/2$.

And [T2] follows from (3a)–(3b):

*Argument 2**: Let n be $|f(a) - f(b)|$ (there is such a number by (3a) & (3b)). Then if $b \succ a$, then $n = f(b) - f(a)$ and $d(a, b) = f(b) - f(a)$. And if these hold, $d(a, b) = n$. Similarly, if $a \succ b$, then $n = f(a) - f(b)$ and $d(a, b) = f(a) - f(b)$. And if these hold, $d(a, b) = n$.[12]

What does this mean? It means that [T1]–[T2] are consistent and compatible because (0)–(1)

---

[11]Similarly, (1), (3b), and (4) imply the mirror images of [T1] and [T1a]:

[T3] $\exists n \exists m < n([b \succ a \rightarrow d(a, b) = m] \,\&\, [a \succ b \rightarrow d(a, b) = n])$.
[T3a] $\exists n > 0([b \succ a \rightarrow d(a, b) = n] \,\&\, [a \succ b \rightarrow d(a, b) = n])$.

[12]Note that these arguments do not use controversial rules of inference that result from identifying the indicative conditional with the material conditional (see the discussion in section 3). Argument *2** is more elaborate than the first part of Argument *2* (or Smullyan's argument for Proposition *2*), and shows that it is not necessary to use such rules.

are so. They are satisfied by any possible situation (e.g., *S*) that satisfies (0)–(1). But no possible situation can satisfy both [P1] and [P2], which are inconsistent. So [T1] and [T2] do not imply [P1] and [P2]. Moreover, [T1]–[T2], taken together, contradict both [P1] and [P2].

To see this, note that [P1] and [P2] share two definite descriptions. So both of them imply the *adequacy conditions* for them:

[A1] $\exists m([b \succ a \rightarrow d(a, b) = m] \ \&\ \forall k([b \succ a \rightarrow d(a, b) = k] \rightarrow k = m))$.

[A2] $\exists m([a \succ b \rightarrow d(a, b) = m] \ \&\ \forall k([a \succ b \rightarrow d(a, b) = k] \rightarrow k = m))$.

[T1]–[T2] contradict both [P1] and [P2], because they contradict the conjunction of [A1] and [A2]. To see this, suppose that they do not. Then the following instances of [T1]–[T2] and [A1]–[A2] must all be jointly consistent:

$[b \succ a \rightarrow d(a, b) = n] \ \&\ [a \succ b \rightarrow d(a, b) = m] \ \&\ m < n$.

$[b \succ a \rightarrow d(a, b) = n_1] \ \&\ [a \succ b \rightarrow d(a, b) = n_1]$.

$[b \succ a \rightarrow d(a, b) = m_1] \ \&\ \forall k([b \succ a \rightarrow d(a, b) = k] \rightarrow k = m_1)$.

$[a \succ b \rightarrow d(a, b) = m_2] \ \&\ \forall k([a \succ b \rightarrow d(a, b) = k] \rightarrow k = m_2)$.

But these imply ‘$n = m_1$’, ‘$n_1 = m_1$’, ‘$m = m_2$’, and ‘$n_1 = m_2$’, which contradict ‘$m < n$’ because they imply ‘$m = n$.’

[T1]–[T2], we have seen, contradict both [P1] and [P2]. So the attempt to show that a possible situation satisfies the latter by showing that it satisfies the former is self-defeating. To succeed in the first half of the attempt is to doom its second half.

In particular, any possible situation that satisfies (0)–(1) (e.g., S or the Smullyan situation) must violate both [P1] and [P2], because it must satisfy both [T1] and [T2]. It might be useful to illustrate this with an example. (0)–(1) imply '$b \succ a \vee a \succ b$.' So a situation that satisfies (0)–(1) (e.g., S) must satisfy either '$b \succ a$' or '$a \succ b$.' Now, let S_a^{10} be a possible situation that satisfies '$b \succ a$' and '$f(a) = 10$' as well as (0)–(1). Then it must satisfy both of the following conditions:[13]

(a) $a \succ b \rightarrow d(a, b) = 10/2.$

(b) $a \succ b \rightarrow d(a, b) = 10.$

These imply the negation of [A2].[14] So S_a^{10} violates both [P1] and [P2]. It is the same with any possible situation that satisfies (0)–(1) and '$b \succ a$.' Similarly, any possible situation that satisfies (0)–(1) and '$a \succ b$' violates [P1] and [P2] because it violates the other adequacy condition, [A1].

Both Argument *1* and Argument *2* fail, we have seen, because their second parts break down for the same reason. It is straightforward to see that the same analysis applies to Smullyan's original

---

[13]We can see that S_a^{10} satisfies (a) and (b) by applying the second parts of Arguments *1** & *2**, because (a) and (b) are the second conjuncts of instances of [T1a] and [T2]:

$[b \succ a \rightarrow d(a, b) = 10]$ & $[a \succ b \rightarrow d(a, b) = 10/2]$.
$[b \succ a \rightarrow d(a, b) = 10]$ & $[a \succ b \rightarrow d(a, b) = 10]$.

See Arguments *1a* & *2a* given in section 3.

[14](a)–(b) imply the negation of '$\forall k([a \succ b \rightarrow d(a, b) = k] \rightarrow k = m)$'; this together with (a)–(b) implies '$10/2 = m$' and '$10 = m$.'

arguments that lead to Propositions *1* & *2*. The descriptions of the Smullyan situation imply the Smullyan analogues of [T1] and [T2]:

[T1'] There is a number n, and a number m smaller than n such that if you gain on the trade, you will gain \$$n$, whereas if you lose on the trade, you will lose \$$m$.

[T2'] There is a number n such that if you gain on the trade, you will gain \$$n$, and that if you lose on the trade, you will lose \$$n$.

These theses, taken together, contradict both propositions, just as [T1]–[T2] contradict both [P1] and [P2].[15] So any possible situation that satisfies the descriptions (e.g., the Smullyan situation) must violate the propositions while satisfying [T1'] and [T2']. Moreover, the attempt to show that a possible situation satisfies the former by showing that they satisfy the latter is self-defeating.

3. Compatibility of Contrary Conditionals

Smullyan's paradox helps to highlight an interesting feature of indicative conditionals. Say that two

---

[15]We can capture the Smullyan situation by adding to (0)–(1) the following conditions:

You will gain on the trade if and only if Baba has more than Ali.
You will lose on the trade if and only if Ali has more than Baba.
You will gain \$$n$ on the trade if and only if Baba has more than Ali and $d(a, \mathrm{b}) = n$.
You will lose \$$n$ on the trade if and only if Ali has more than Baba and $d(a, \mathrm{b}) = n$.

Given these conditions, [T1'] and [T2'] are equivalent to [T1] and [T2], respectively; the former result essentially from replacing the right sides of the conditions in the latter with their left sides. Similarly, given the conditions, Propositions *1* and *2* are equivalent to [P1] and [P2], respectively.

(indicative) conditionals are *contrary*, if they have the same antecedent but incompatible consequents; and *strongly contrary*, if they are contrary and their antecedent is satisfiable.[16] Then (a) and (b), for example, are strongly contrary indicative conditionals (for there is a possible situation that satisfies '$a \succ b$'). It is implausible to hold that any contrary conditionals are incompatible. Two contrary conditionals whose common antecedent is a contradiction (e.g., '$A \;\&\; {\sim}A \rightarrow A$' and '$A \;\&\; {\sim}A \rightarrow {\sim}A$') might both hold. But some might think it is plausible to hold a weaker thesis:

[C] Strongly contrary indicative conditionals are incompatible.[17]

Ramsey, for example, suggests that "in a sense 'If *p*, *q*' and [its strict contrary][18] are contradictories" in cases where the negation of *p* is not believed to be certain (1929, 247). And Stalnaker holds "the denial of a conditional is equivalent to a conditional with the same antecedent and opposite consequent (provided that the antecedent is not impossible)" (1968, 49).[19] The apparent plausibility of their view, which implies [C], might explain the seeming intuitive force of the second parts of

---

[16] I say a statement is *satisfiable* (on a situation or possible world *w*), if there is a possible situation (accessible from *w*) that satisfies it. So ϕ is satisfiable (on *w*) if and only if $\diamond\phi$ holds (on *w*).

[17] That is, $\diamond\phi, (\phi \rightarrow \psi), {\sim}\diamond(\psi \;\&\; \chi) \vDash {\sim}(\phi \rightarrow \chi)$ or, equivalently, $\diamond\phi, (\phi \mathrel{\Box\!\!\rightarrow} \psi), (\phi \mathrel{\Box\!\!\rightarrow} \chi) \vDash \diamond(\psi \;\&\; \chi)$. Stalnaker (1981) gives an axiom of conditional logic that amounts to [C]. See note 19.

[18] I.e., 'If *p*, ~*q*' (instead of '~', Ramsey uses the raised bar '–').

[19] This idea is captured by Axiom (a2) of his formal system C2 of conditional logic (*ibid*., p. 48). Harper calls it "Stalnaker's axiom" (1981, 6). Because Stalnaker does not distinguish the indicative conditional from the counterfactual conditional, the axiom may be taken to have two versions. See the last paragraph of this section.

Smullyan's argument for Propositions *1* & *2*. In presenting his argument for Proposition *1*, for example, Smullyan does not seem to find it necessary to elaborate on the second part, namely, the last step of the argument indicated by the italicized *then*:

> Then if you gain on the trade, you will gain n dollars, but if you lose on the trade, you will lose n/2 dollars. . . . *then* the amount you gain, if you do gain . . . is greater than the amount you will lose, if you do lose . . . . (1992, 191; my italics)

Some might invoke [C] to justify the step.

To see this, too, it is useful to examine the second parts of Arguments *1* & *2*. Because '$a \succ b$' is satisfiable, [C] implies that any situation satisfies the following conditions:[20]

(γ) $\forall k \forall l([a \succ b \rightarrow d(a, b) = k] \;\&\; [a \succ b \rightarrow d(a, b) = l] \rightarrow k = l)$.

This implies the equivalence of the following conditions:[21]

(α) $a \succ b \rightarrow d(a, b) = n$. (If Ali has more than Baba, *the difference between the amounts in Ali and Baba* is \$*n*.)

(β) $(\iota k)[a \succ b \rightarrow d(a, b) = k] = n$. (*The difference between the amounts in Ali and Baba*

[20]Suppose '$\diamond\, a \succ b$', '$[a \succ b \rightarrow d(a, b) = k]$', and '$[a \succ b \rightarrow d(a, b) = l]$' hold. Then, by [C], '$d(a, b) = k$' and '$d(a, b) = l$' are compatible. This implies '$\diamond\, k = l$', which implies '$k = l$.'

[21](β) is equivalent to the conjunction of (α) and (γ).

in case Ali has more than Baba is \$*n*.)

Similarly, [C] implies the equivalence of their siblings:

(α') $b \succ a \rightarrow d(a, b) = n$

(β') $(\iota k)[b \succ a \rightarrow d(a, b) = k] = n$.

Using those equivalences, one can derive [P1] from [T1], and [P2] from [T2].[22]

This derivation of [P1]–[P2] from [T1]–[T2] that rests on [C], however, does not yield a justification of the second parts of Arguments *1* & *2*. Instead, it shows that [C] is false. For there are possible situations that satisfy [T1] while violating [P1], as we have seen.

And we can directly see that [C] is false. S_a^{10}, for example, satisfies both (a) and (b), and witnesses the falsity of [C]. Similarly, any possible situation that satisfies (0)–(1) witnesses its falsity. Those among them that satisfy '$b \succ a$' (e.g., S_a^{10}) satisfy strongly contrary substitution instances of (α) (e.g., (a)–(b)); and those that satisfy '$a \succ b$' those of its sibling, (α').[23]

This should be no surprise to those who identify the *indicative* conditional (in symbols, '$\rightarrow$') as the *material* conditional (in symbols, '$\supset$'). Because material conditionals hold as long as their antecedents fail, any possible situation witnesses the falsity of the material conditional cousin of [C]:

[22]Given the equivalences, [P1] is equivalent to [T1], and [P2] to [T2].

[23]By *substitution instances* of (α), I mean sentences that one can obtain from them by replacing the variable '*n*' with numerals.

[C’] Strongly contrary material conditionals are incompatible.

But it is controversial whether the indicative conditional can be identified with the material conditional.[24] So some might attempt to revive Smullyan’s paradox by holding that [C], unlike [C’], is true by distinguishing the indicative from the material. But my analysis of the paradox, including the reasoning that leads to counterexamples to [C], does not rest on identifying them.

Note that any situation that satisfies ‘$b \succ a$’ (e.g., S_a^{10}) satisfies all the following material conditionals:

(a’) $a \succ b \supset d(a, b) = 10/2$.

(b’) $a \succ b \supset d(a, b) = 10$.

(c) $a \succ b \supset d(a, b) = 7$.

(d) $a \succ b \supset d(a, b) = \pi$.

To show this, it is necessary to invoke features specific to material conditionals that license, e.g., the following implications:

[M1] $\sim\phi \vDash (\phi \supset \psi)$.

[M2] $\psi \vDash (\phi \supset \psi)$.

[24] For a discussion of reasons against identifying them, see, e.g., Priest (2008, esp. section 1.9). See also Priest (2002, 15).

But it is not clear whether the indicative conditional is subject to the cousins of these implications. So defenders of [C] might deny that S_a^{10}, for example, satisfies all the indicative cousins of (a'), (b'), (c), (d), etc. I think one might plausibly deny that the situation satisfies the indicative cousins of (c) & (d), but this does not mean that it is the same with (a) & (b), the indicative cousins of (a') & (b'). Both (a) and (b) hold on S_a^{10}, because '$f(a) = 10$' and '$b \succ a$', given (0)–(1), imply them. We can show this by applying the second parts of Arguments *1* & *2* as follows:

> *Argument 1a*: Let $f(a)$ be 10. Then if $a \succ b$, $f(b) = 10/2$ (by (1) & (4)). And if $a \succ b$ and $f(b) = 10/2$, then $d(a, b) = f(b) - f(a) = 10 - 10/2 = 10/2$.

> *Argument 2a*: Suppose that $f(a) = 10$ and $b \succ a$. And let $n = |f(a) - f(b)|$. Then if $a \succ b$, then $n = f(a) - f(b)$ and $d(a, b) = f(a) - f(b)$. And if these hold, $d(a, b) = n$. And $n = |f(a) - 2{\cdot}f(a)| = 10$ (for $f(b) = 2{\cdot}f(a)$ by (4)). So if $a \succ b$, $d(a, b) = 10$.

These arguments, note, do not invoke the indicative cousins of [M1]–[M2] or any other controversial implications that result from identifying the indicative conditional with the material. Even those who reject the identification would take all the inference rules used there[25] to be natural and legitimate. Those are rules that, unlike [M1]–[M2], must be retained in any adequate analysis of the indicative conditional. If so, one cannot defend [C] by rejecting some of them.

[25]E.g., (i) $A \rightarrow B, B \rightarrow C \vdash A \rightarrow C$; (ii) $A \rightarrow B \vdash A \rightarrow (A \& B)$; and (iii) $A \rightarrow B, A \rightarrow C \vdash A \rightarrow (B \& C)$. Stalnaker (1968, 48) rejects the transitivity (i), but the counterexamples he gives concerns only the counterfactual cousin of (i), which I agree is unsound. Instead of (i), in any case, one can use a variant of (i): (i') $A \rightarrow B, (A\&B) \rightarrow C \vdash A \rightarrow C$. Note that the counterfactual cousin of (i') is also sound.

One can show that any possible situations that satisfy conditions (0)–(1) are counterexamples to [C], we have seen, using only a meager set of natural inference rules that pertain to the indicative conditional. I think this, unlike the falsity of [C'], is somewhat surprising. It is especially so considering that the conditions are very weak.[26] They boil down to essentially one condition, (0).[27] This means that one can create a situation that satisfies strong contraries simply by putting, for whatever reason, twice as much money in one envelope than another![28]

Smullyan's paradox, we have seen, calls for clear recognition of compatibility of strongly contrary indicative conditionals. So does a well-known paradox in decision theory, Newcomb's.

Suppose that there are two boxes, *A* and *B*, and that you have a choice between the content of just box *A* (choice α) and the contents of both boxes (choice β). Now, *B* has \$10, while *A* has \$100 if and only if you will take choice α, and \$0 if and only if you will take choice β. (These two conditionals hold because the omniscient God put \$100 or none in *A* according to his perfectly accurate prediction about which choice you will take, or for some other reason. It does not matter what the reason is.) Using '$T(x)$' and '$G(n)$' for 'You will take choice *x*' and 'You will gain \$*n*', we can formulate two conditions that hold on the Newcomb situation described above, *N*, as follows:

(i) $[T(\alpha) \rightarrow G(100)]\ \&\ [T(\beta) \rightarrow G(10)]$.

(ii) $([T(\alpha) \rightarrow G(0)]\ \&\ [T(\beta) \rightarrow G(0{+}10)]) \vee ([T(\alpha) \rightarrow \mathrm{G}(100)]\ \&\ [T(\beta) \rightarrow G(100{+}10)])$.

---

[26]Note also that no conditional is used in their formulations.

[27]See note 8.

[28]Moreover, we can replace the numeral '2' in (0) with the numeral for any number greater than 1, and run essentially the same arguments. This means that any possible situation where two envelopes have two different positive amounts satisfies strongly contrary indicatives.

Condition (i) follows from the above biconditionals; (ii) holds because B has either \$0 or \$100. Now, one might take, e.g., the conditional '[$T(\alpha) \rightarrow G(100)$]' to amount to the corresponding statement about the result of taking a choice, 'The result of your taking choice α is your gaining \$100', and accept the following principle of choice:

[PC] $[T(x) \rightarrow G(m)]$ & $[T(y) \rightarrow G(n)]$ & $m > n \rightarrow B(x, y)$, where '$B(x, y)$' is for '$x$ is a better choice than y.'

Using this principle, one might invoke (i) to conclude that α is a better choice than β. Using the same principle, however, one might invoke (ii) to draw the opposite conclusion: β is a better choice than α. The results is a version of Newcomb's paradox that invokes no probability.[29]

The solution to this paradox lies in rejecting [PC], as I argue in Yi (2003). The apparent plausibility of [PC] vanishes, I think, if one realizes that the Newcomb situation N satisfies strongly incompatible indicatives germane to the results of taking the two choices. To see this, note that it must satisfy one of two disjuncts of (ii). Now, if it satisfies the left disjunct, it satisfies both '[$T(\alpha) \rightarrow G(100)$]' and '[$T(\alpha) \rightarrow G(0)$]'; and both '[$T(\beta) \rightarrow G(10)$]' and '[$T(\beta) \rightarrow G(100+10)$]' otherwise. If so, the two incompatible conditionals that it satisfies (e.g., '[$T(\alpha) \rightarrow G(100)$]' & '[$T(\alpha) \rightarrow G(0)$]') cannot both be taken to indicate the results of taking the relevant choice (e.g., α). To see this is to see that [PC] is not a plausible principle of choice at all. It must be distinguished from the

[29]This elegant version of Newcomb's paradox is due to Priest (2002), who uses it to conclude that the Newcomb situation N is a rational dilemma, a situation in which rationality requires one to do incompatible things (e.g., taking choice α and taking choice β). In Yi (2003), I argue that Priest's argument rests on a wrong principle of rational choice, a version of [PC], that results from ignoring the falsity of [C].

following:

[PC'] $(\iota m)[T(x) \rightarrow G(m)] > (\iota n)[T(y) \rightarrow G(n)] \rightarrow B(x, y)$.

[PC"] A choice is better than another, if the result of taking the first choice is gaining \$*m* while the result of taking the second is gaining \$*n* and *m* is greater than *n*.

I think these are plausible principles, but one cannot derive [PC] from these without invoking a wrong principle, such as [C]. Now, recognizing the falsity of [C] is crucial to discerning the gap between these plausible principles and [PC]. The recognition provides the key to the solution of Newcomb's paradox, just as it is the lynchpin of the solution to Smullyan's.

Now, the solutions lead to rejection of many popular accounts of conditionals. Ramsey and Stalnaker, for example, accept [C], as noted above. It is a central component of their accounts of acceptability or truth of conditionals. Ramsey suggests that "If two people are arguing 'If *p* will *q*?' and are both in doubt as to *p*, they are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q*" (1929, 247). Stalnaker turns the suggestion about deciding whether or not to accept conditionals to an account of their truth conditions. He gives an informal exposition of the account as follows:

> Consider a possible world in which *A* is true, and which otherwise differs minimally from the actual world. *"If A, then B" is true (false) just in case B is true (false) in that possible world.* (1968, 45)

The same person cannot accept two strongly contrary conditionals at the same time on Ramsey's account; similarly, two such conditionals cannot both be true on Stalnaker's. Ramsey and Stalnaker do not clearly distinguish the indicative conditional from the counterfactual. If one distinguishes them, as I think one should,[30] their accounts can be taken to have two versions: one for the indicative, and one for the counterfactual. So Stalnaker's statement, quoted above, to the effect that two strongly contrary conditionals are incompatible can be taken to have two versions, [C] and its counterfactual cousin:

[C*] Strongly contrary counterfactual conditionals are incompatible.[31]

In defending the statement, Stalnaker gives only a defense of [C*], noting the observation, made by "Goodman and Chisholm in their early papers on counterfactuals, that the normal way to contradict a counterfactual is to contradict the consequent, keeping the same antecedent" (*ibid.*, 49). I think Stalnaker's account yields the correct about [C*],[32] but we have seen a good reason to conclude that

---

[30]They have different logics. One difference between them is that subsitutivity of identity holds in the indicative context, but not in the counterfactual context. Suppose that <u>Ali's amount</u> is <u>$20</u> (and Baba's $10), and that <u>Ali's amount</u> would have been less than $10 if Ali had less than Baba, and yet one cannot conclude that <u>$20</u> would have been less than $10 if Ali had less than Baba. (Both suppositions hold on the mirror image of the situation S_1 presented in section 4. See note 47.) And the counterfactual cousin of [C], [C*] below, holds while [C] fails.

[31]That is, $\Diamond\phi, (\phi \mathbin{\Box\!\!\rightarrow} \psi), {\sim}\Diamond(\psi \,\&\, \chi) \vDash {\sim}(\phi \mathbin{\Box\!\!\rightarrow} \chi)$ or, equivalently, $\Diamond\phi, (\phi \mathbin{\Box\!\!\rightarrow} \psi), (\phi \mathbin{\Box\!\!\rightarrow} \chi) \vDash \Diamond(\psi \,\&\, \chi)$. Lewis (1973, 16) gives a special case of the thesis: $\Diamond\phi, (\phi \mathbin{\Box\!\!\rightarrow} \psi) \vDash {\sim}(\phi \mathbin{\Box\!\!\rightarrow} {\sim}\psi)$. And this amounts to Axiom (a3), "Stalnaker's axiom", in Stalnaker's system C2 of conditional logic, if the system is taken to concern counterfactuals. Note that the special case implies the general case.

[32]I think the truth of [C*] is intuitively clear, which explains the observation made by Goodman and Chisholm. It might be useful to see how it results from the standard possible world analyses of counterfactuals given by Stalnaker and Lewis. Say that a possible world is a *ϕ-world*, if it satisfies ϕ; and that a class, C, of possible worlds is a *ϕ-neighborhood of* a possible world, w,

the account yields the wrong result about its indicative cousin, [C].[33]

4. Counterfactuals and Smullyan's Paradox

In his analysis (2002) of Smullyan's paradox, Chase holds that the conditional used to formulate the paradox must be considered the counterfactual conditional. He argues that the two conditionals in [T1] (or its Smullyan analogue, [T1']) must be counterfactuals. Their antecedents, he says, "cannot be true together, so at least one is contrary to fact. Since we do not know which is, both conditionals are counterfactual" (*ibid*., 158). Although their antecedents cannot both be true, as he notes, this does not mean that "there is no way they can both be . . . interpreted" as indicative conditionals (*ibid*., 160); nor does it mean that [T1] (or [T1']), so interpreted, must be false or fail to hold on *S* (or the Smullyan situation). Consider, for example, 'There are integers *m* and *n* such that $\chi_R(\pi) = m$ if π is a rational number, and $\chi_R(\pi) = n$ if π is not a rational number, and $m > n$', where 'χ_R' is for the characteristic function for the property of being a rational number. Surely, both conditionals in

---

if any member of *C* is a ϕ-world accessible from *w*, and is closer to *w* than is any ϕ-world that is not a member of *C*. Then the following holds according to the standard analyses:

> [*] If a possible world, *w*, satisfies both $\Diamond\phi$ and $(\phi \mathbin{\Box\!\!\rightarrow} \psi)$, there is a ϕ-neighborhood of *w* all the member of which are ψ-worlds.

And note that if C_1 and C_2 are ϕ-neighborhoods of the same world, C_1 and C_2 overlap (in fact, C_1 includes C_2 or vice versa). Now, let *w* satisfies $\Diamond\phi$, $(\phi \mathbin{\Box\!\!\rightarrow} \psi)$, and $(\phi \mathbin{\Box\!\!\rightarrow} \chi)$. Then let C_1 be a ϕ-neighborhood of *w* all the member of which are ψ-worlds, and C_2 a ϕ-neighborhood of *w* all the member of which are χ-worlds. (There are such classes by [*].) And let *w** be a member of both C_1 and C_2. Then it is accessible from *w*, and satisfies both ψ and χ. So $\Diamond (\psi \,\&\, \chi)$ holds.

[33] I think the falsity of [C] yields a response to the challenge to modus ponens initiated by McGee (1985). See also Etlin (2009, 682f). I leave it for another occasion to spell out the response.

this sentence can be taken to be indicative; and it is true when they are so taken.[34] If so, there is no reason not to take the conditionals in Smullyan's formulation of his paradox to be indicative. They *are* indicative, not counterfactual, conditionals. And [T1] and [T2] (or their Smullyan analogues), as taken to involve indicative conditionals, hold on *S* (or the Smullyan situation), as we have seen.

But some might attempt to formulate a counterfactual cousin of Smullyan's paradox by taking the conditional used to formulate it to be counterfactual. To do so, they might attempt to show that situation *S* must satisfy the counterfactual cousins of [P1] & [P2] by showing that (0)–(1) imply the counterfactual cousins of [T1] and [T2]:[35]

[T1*] $\exists n \exists m < n([b \succ a \;\Box\!\!\rightarrow d(a, b) = n] \;\&\; [a \succ b \;\Box\!\!\rightarrow d(a, b) = m])$.

[T2*] $\exists n([b \succ a \;\Box\!\!\rightarrow d(a, b) = n] \;\&\; [a \succ b \;\Box\!\!\rightarrow d(a, b) = n])$.

But one cannot show this by modifying Arguments *1* & *2*. One can derive the indicative conditional (α), '$[a \succ b \rightarrow d(a, b) = n]$', from (0)–(1) and '$f(a) = n$',[36] but one cannot derive its counterfactual cousin from them. The argument for (α) must involve substituting 'n' for '$f(a)$' in the following conditional using the identity '$f(a) = n$':

[34]It follows from 'Any real number r is such that $\chi_R(r) = 1$ if r is a rational number, and $\chi_R(\mathrm{r}) = 0$ if r is not a rational number'. See, e.g., Priest (2002, 14f) for more on indicative conditionals with false antecedents.

[35]I use '$\Box\!\!\rightarrow$' for the counterfactual conditional, indicated in English by the subjunctive mood of the antecedent and consequent. '$[a \succ b \;\Box\!\!\rightarrow d(a, b) = n]$', for example, amounts to 'If Ali *had* more than Baba, the difference between the amounts in Ali and Baba *would* have been *n*.'

[36]See Argument *1a* or the second part of *Argument 1**.

$$[a \succ b \rightarrow f(a) = 2 \cdot f(b)].$$

But one cannot use the identity to substitute 'n' for '$f(a)$' in its counterfactual cousin:

$$[a \succ b \mathrel{\Box\!\!\rightarrow} f(a) = 2 \cdot f(b)].$$

Substitutivity of identity fails in the counterfactual context, as noted above.[37] So one cannot turn Argument *1** to an argument for [T1*].[38] Another problem arises from the fact that the descriptions of *S*, (0)–(1), involve no counterfactuals or necessities. The conditions imply (4), for example, but not its counterfactual cousin:

(4*) $\forall x \forall y[x \succ y \mathrel{\Box\!\!\rightarrow} f(x) = 2 \cdot f(y)]$. (If an envelope on the table *had* more than another, the former *would have had* twice as much as the latter.)

But it would be necessary to use this to turn Argument *1** into an argument for [T1*].[39]

To solve the second problem, one might try strengthening the conditions for *S* by adding,

---

[37]See note 30.

[38]It is the same with Argument *2**, which also invokes substitutivity of identity.

[39]Similarly, the descriptions of the Smullyan situation do not imply the Smullyan analogues of [T1*] and [T2*], i.e., the counterfactual cousins of [T1'] and [T2']. Chase (2002) correctly argues that the descriptions do not imply the analogue of [T1*], but his argument does not apply to [T2*]. He goes a step further to claim that the Smullyan situation must fail to satisfy the analogue of [T1*] because "only one of [the conditionals in it] can be true" (*ibid.*, p. 159). This is not correct. Both [T1*] and [T2*], as we shall see, are compatible with (0)–(1). Similarly, their analogues are compatible with the descriptions of the Smullyan situation. See the last two paragraphs of this section.

e.g., (4*).[40] One can then use (4*) to get '[$a \succ b$ □→ $f(a) = 2{\cdot}f(b)$].' But this does not help to derive '[$a \succ b$ □→ $n = 2{\cdot}f(b)$]' using '$f(a) = n$.' To solve this problem, one might assume the *necessary* identity '$\Box\, f(a) = n$.' So one might attempt to present a paradox by imagining a possible situation, S^*, that satisfies the following conditions:[41]

*Descriptions of S**:

(0*) $\exists x \exists y[\forall z(z = x \vee z = y)\ \&\ \Box\,(f(x) = 2{\cdot}f(y)\ \&\ f(y) > 0)]$.

(1) $\exists x\, x{=}a\ \&\ \exists x\, x{=}b\ \&\ a{\neq}b$.

(5) $\exists n \Box\, f(a) = n$.

(6) $\exists n \Box\, d(a, b) = n$.

We can show that (0*), (1), and (5) imply [T1*] by modifying Argument *1**, and that (6) implies [T2*] by modifying Argument *2**.[42] So S^* must satisfy both [T1*] and [T2*]. Now, one might argue that these are incompatible because they imply the counterfactual cousins of [P1] and [P2]:

---

[40]Or, equivalently, one might replace (4) with (4*) in (1)–(4).

[41]The conditions do not include (4*), but (0*) implies it.

[42]Modifying *Argument 1** yields the following, which implies [T1*]:

[T1a*] $\exists n > 0$([$b \succ a$ □→ $d(a, b) = n$] & [$a \succ b$ □→ $d(a, b) = n/2$]).

To modify Arguments *1* & *2*, it is necessary to invoke the counterfactual cousins of the rules of inference used in them (e.g., those mentioned in note 25), e.g., (i*) A □→ B, (A & B) □→ C ⊢ A □→ C, (ii*) A □→ B ⊢ A □→ (A & B), and (iii*) A □→ B, A □→ C ⊢ A □→ (B & C). Note that (i*) differs from the straightforward modification of (i): A □→ B, B □→ C ⊢ A □→ C. This rule is unsound. But we can use (i*) instead.

$[\text{P1*}]\quad (\iota k)[b \succ a \;\Box\!\!\rightarrow d(a, b) = k] > (\iota k)[a \succ b \;\Box\!\!\rightarrow d(a, b) = k]).$

$[\text{P2*}]\quad (\iota k)[b \succ a \;\Box\!\!\rightarrow d(a, b) = k] = (\iota k)[a \succ b \;\Box\!\!\rightarrow d(a, b) = k]).$

The problem with this argument is that [T1*]–[T2*] do not imply [P1*]–[P2*] just as [T1]–[T2] do not imply [P1]–[P2]. Let S_a^{10}* be a situation that satisfies (0*), (1), '$\Box f(a) = 10$', and '$\Box\, b \succ a$.' Then it satisfies [T1*]–[T2*], because it satisfies the following conditions:[43]

(a*) $\quad a \succ b \;\Box\!\!\rightarrow d(a, b) = 10/2.$

(b*) $\quad a \succ b \;\Box\!\!\rightarrow d(a, b) = 10.$

(e) $\quad b \succ a \;\Box\!\!\rightarrow d(a, b) = 10.$

But it satisfies neither [P1*] nor [P2*]. Both of these imply the adequacy condition for the definite description '$(\iota k)[a \succ b \;\Box\!\!\rightarrow d(a, b) = k])$':

$[\text{A2*}]\quad \exists m([a \succ b \;\Box\!\!\rightarrow d(a, b) = m] \;\&\; \forall k([a \succ b \;\Box\!\!\rightarrow d(a, b) = k] \rightarrow k = m)).$

The negation of this follows from (a*)–(b*), which imply the negation of '$\exists m \forall k([a \succ b \;\Box\!\!\rightarrow d(a, b)$

---

[43]One can show that (a*) and (b*) hold on S_a^{10}* by deriving them from the descriptions of S_a^{10}* by modifying Arguments *1a* & *2a* as indicated in note 41. ((a*) and (b*), though contraries, can both hold on S_a^{10}*, because S_a^{10}* satisfies '$\sim\Diamond\, a \succ b$.') Note that on the standard possible-word analyses, counterfactuals are subject to the modal-counterfactual versions of [M1] and [M2]: $\Box\sim\phi \vDash (\phi \;\Box\!\!\rightarrow \psi)$, and $\Box\psi \vDash (\phi \;\Box\!\!\rightarrow \psi)$. For the standard analyses, see, e.g., Stalnaker (1968) and Lewis (1973, 16). But I do not assume the correctness of these results, and the suggested arguments for (a*) and (b*) do not invoke them. So one cannot apply the arguments to derive, e.g., '$a \succ b \;\Box\!\!\rightarrow d(a, b) = \pi$.'

$= k] \to k = m$.'

The above counterexample, S_a^{10}*, violates '$\Diamond\, a \succ b$.' Its mirror image, which violates '$\Diamond\, b \succ a$', is also a counterexample. Now, one might eliminate the counterexamples by adding the following condition:

(7) $\Diamond\, b \succ a \ \&\ \Diamond\, a \succ b$.

Given this, [T1*] and [T2*] imply [P1*] and [P2*], respectively. For [C*], "Strongly contrary counterfactual conditionals are incompatible", is true, as noted in section 3. Just as its indicative cousin, [C], given (7), implies the equivalence of (α) and (β), so does [C*], given (7), imply the equivalence of their counterfactual cousins:[44]

(α*) $a \succ b \mathrel{\Box\!\!\rightarrow} d(a, b) = n$.

(β*) $(\iota k)[a \succ b \mathrel{\Box\!\!\rightarrow} d(a, b) = k] = n$.

But this leads to no paradox. For it would be wrong to assume that there is a possible situation that satisfies (7) as well as the descriptions of S*. There is no such situation because the descriptions contradict (7). They imply the following:

(8) $\exists n(\Box\, [f(a) = n \ \&\ f(b) = 2n] \vee \Box\, [f(a) = n \ \&\ f(b) = n/2])$.

---

[44]For [C*], given '$\Diamond\, a \succ b$', yields the counterfactual cousin of (γ): $\forall k \forall l([a \succ b \mathrel{\Box\!\!\rightarrow} d(a, b) = k] \ \&\ [a \succ b \mathrel{\Box\!\!\rightarrow} d(a, b) = l] \to k = l)$. (Note that the last conditional sign in this sentence is the indicative conditional.) See note 20.

And this contradicts (7).

Moreover, (8) is equivalent to the conjunction of [T1*] and [T2*], as we can see using [C*].[45] But (8), which contradicts (7), contradicts both [P1*] and [P2*]. So both of these must fail to hold on any possible situation that satisfies [T1*] and [T2*].[46] This means that the attempt to formulate a counterfactual version of Smullyan's paradox, like the attempt to revive the indicative version, is self-defeating.

This completes my analysis of the counterfactual version of Smullyan's paradox. The analysis, note, draws a parallel to the analysis of its indicative version given in section *2*. But we can see an interesting disparity between the two versions if we begin by assuming that situation S satisfies (7), that is, by considering a possible situation, $S^{\dagger}$, that satisfies (7) as well as (0)–(1). The additional assumption makes no difference to the analysis of the indicative paradox. Even assuming (7), [T1], for example, fails to imply [P1]. For the indicative conditional (α) is not equivalent to the corresponding identity (β), even assuming satisfiability of their antecedent, '$a \succ b$.' The reason is that [C] is false. But its counterfactual cousin, [C*], is true, and this makes it different with the counterfactual cousins of (α) and (β), (α*) and (β*). These are equivalent, if '$a \succ b$' is satisfiable. Accordingly, [T1*] and [T2*], given (7), are equivalent to [P1*] and [P2*], respectively, as we have seen.

---

[45] Assuming the negation of (7), (8) is equivalent to the conjunction of [T1*] and [T2*]. And (8) implies the negation of (7), and so does the conjunction of [T1*] and [T2*]. (To see this, suppose that '$[b \succ a \;\Box\!\!\rightarrow d(a, b) = n]$ & $[a \succ b \;\Box\!\!\rightarrow d(a, b) = m]$', '$m < n$', and '$[b \succ a \;\Box\!\!\rightarrow d(a, b) = k]$ & $[a \succ b \;\Box\!\!\rightarrow d(a, b) = k]$', as well as (7) hold. Then '$n = d(a, b) = k$' and '$m = d(a, b) = k$' hold (by [C*]). These imply '$n = m$', which contradicts '$m < n$.')

[46] Similarly, any situation that satisfies the Smullyan analogues of [T1*] and [T2*] must violate both of the counterfactual cousins of Propositions *1* & *2*.

If so, one might ask whether $S^\dagger$ satisfies [T1*] or [T2*]. It cannot satisfy both, because they, given (7), are incompatible, as we have seen. If so, does it satisfies one of them? And if so, which one? The answer is that one cannot tell without more information about $S^\dagger$ because there are various possible situations that satisfy (7) as well as (0)–(1). Some of them satisfy [T1*] (and so [P1*]), some others [T2*] (and so [P2*]), and yet others neither. It would be useful to examine representative samples of the various situations.

Let S_1, S_2, and S_3 be possible situations that satisfy (7) as well as (0)–(1). In all of them, suppose, Ali turns out to have \$10, and Baba \$20. But the amounts have been determined in different ways:

S_1: The amount in Ali was fixed as \$10, but the amount in Baba was determined by tossing a fair coin: \$5 for heads and \$20 for tails. (The coin fell tails.)

S_2: The amounts in Ali and Baba were fixed as either (\$10, \$20), viz., \$10 for Ali and \$20 for Baba, or (\$20, \$10). One decided which of the two combinations of amounts to put in Ali and Baba with one toss of a fair coin: the former combination for heads, the latter for tails. (The coin fell heads.)

S_3: The amounts in Ali and Baba were decided with two tosses of a fair coin: the first for Ali's amount (\$10 for heads, \$40 for tails), and the second for Baba's amount (\$5 for heads, \$20 for tails). (The outcomes were heads for Ali, and tails for Baba.)

Now, S_1 satisfies [T1*]; it satisfies (0*) and (5). S_2 satisfies [T2*]; it satisfies (6*). And S_3 satisfies neither.[47] It is a situation in which Ali, if he had more than Baba, might have had \$40 (with Baba still with \$20)–so the difference between their amounts might have been \$20, not \$5 or \$10.

This diversity of the situations that satisfy (0)–(1) & (7) conflicts with Chase's (2002) analysis of Smullyan's paradox, according to which the Smullyan situation (or any possible situation that satisfies its descriptions) must violate the analogue of [T1*]. Chase might object that the Smullyan situation differs from $S^{\dagger}$ in that its conditions include the condition that you choose one of the two envelopes, Ali and Baba, *after* the amounts in them have been determined. He might argue that this means that the Smullyan situation must violate the analogue of [T1*] while satisfying the analogue of [T2*]. Suppose that you actually picked up Ali, with \$10, and that it has less than Baba, with \$20. Then if the envelope you hold had more than the other, you would have had picked up Baba (with Ali and Baba still having \$10 and \$20, respectively) so that the difference between the amounts in them would have been the same, \$10. By amplifying the descriptions of S_1, however, we can specify a possible situation that satisfies the descriptions of the Smullyan situation (as well as (7)) while satisfying the analogue of [T1*] as well. Suppose that although the amounts in Ali and Baba (\$10 and \$20) were determined as in S_1 before you make a choice between the envelopes, you were somehow pre-determined to choose Ali.[48] Then if you would gain on the trade, you would gain \$10; you would then be trading Ali's \$10 for Baba's \$20. If you would lose on the trade, however,

---

[47]The mirror image of S_1, where the amount in Baba is fixed, is also a situation that satisfies neither [T1*] nor [T2*], but it satisfies the mirror image of [T1*]:

$$[\text{T3*}] \quad \exists n \exists m < n([b \succ a \;\Box\!\!\rightarrow d(a, b) = m] \;\&\; [a \succ b \;\Box\!\!\rightarrow d(a, b) = n]).$$

S_3 violates this condition as well.

[48]For example, you might have unknown inclination for envelopes of a certain kind that makes you choose Ali.

you would lose $5; the reason that you would lose would be that Baba would have had $5 because the coin had fallen heads.

References

Broome, J. (1995), “The two-envelope paradox”, *Analysis,* 55(1): 6-11.

Chalmers, D. J. (2002), “The St. Petersburg two-envelope paradox”, *Analysis*, 62(2): 155-157.

Chase, J. (2002), “The non-probabilistic two envelope paradox”, *Analysis* 62(2): 157-60.

Etlin, D. (2009), “The problem of noncounterfactual conditional”, *Philosophy of Science*, 76: 676-688.

Harper, W. L. (1981), “A sketch of some recent developments in the theory of conditionals”, in W. L. Harper, R. Stalnaker, and G. Pearce (eds.), *Ifs: Conditionals, Belief, Chance, and Time* (Dordrecht: Reidel, 1981).

Jackson, F., Menzies, P., and Oppy, G. (1994), “The two envelope ‘paradox’”, *Analysis*, 54(1): 43-45.

Lewis, D. (1973), *Counterfactuals* (Cambridge, Harvard University Press).

McGee, V. (1989), “A counterexample to modus ponens”, *Journal of Philosophy*, 82: 462-471.

Nalebuff, B. (1988), “Cider in your ear, continuing dilemma, the last shall be first, and more”, *Journal of Economic Perspectives,* 2(2): 149-156.

Nalebuff, B. (1989), “Puzzles: the other person’s envelope is always greener”, *Journal of Economic Perspectives*, 3(1): 171-181.

Priest, G. (2002), “Rational dilemmas”, *Analysis* 62(1): 11-16.

Priest, G. (2008), *An Introduction to Non-Classical Logi*c, 2nd ed. (Cambridge: Cambridge University Press).

Priest, G. and Restall, G. (2008), “Envelopes and indifference”, in C. Dégremont, L. Keiff, and H. Rückert (eds.), *Dialogues, Logics and Other Strange Things: Essays in Honour of Shahid Rahman* (London: College Publications, 2008), pp. 277-284.

Ramsey, F. R. (1929), “General propositions and causality”, in Ramsey, *The Foundations of Mathematics and other Logical Essays* (London: RKP, 1931), pp. 237-255.

Smullyan, R. (1992), *Satan, Cantor, and Infinity* (New York: Alfred A. Knopf).

Stalnaker, R. C. (1968), “A theory of conditionals”, *Studies in Logical Theory*, *American Philosophical Quarterly Monograph Series*, No. 2 (Oxford: Blackwell, 1968); reprinted in W. L. Harper, R. Stalnaker, and G. Pearce (eds.), *Ifs: Conditionals, Belief, Chance, and Time* (Dordrecht: Reidel, 1981).

Yi, B.-U. (2003), “Newcomb’s paradox and Priest’s principle of rational choice”, *Analysis* 63(3): 237-42.

Yi, B.-U. (2009), “The two-envelope paradox and causal direction”, unpublished manuscript.