Abstract: This article presents a solution to Smullyan’s version of the two envelope paradox. In doing so, it argues against the probabilistic and possible world approaches to conditionals. These approaches yield the thesis that contrary indicative conditionals whose common antecedent state a possibility (e.g., ‘If Ali has more than Baba, the difference between the amounts in them is $5’ and ‘If Ali has more than Baba, the difference between the amounts in them is $10’) are incompatible. The article argues that this thesis is false although its counterfactual cousin is true. And it explains this disparity between indicative and counterfactual conditionals by clarifying important logical differences between them: (i) substitutivity of identity holds for indicatives but fails for counterfactuals, and (ii) counterfactuals preserve possibility while indicatives preserve only truth.

Consider two contrary conditionals about two envelopes, Ali and Baba:

(a) If Ali has more money than Baba, the difference between the amounts in them is $5.
(b) If Ali has more money than Baba, the difference between the amounts in them is $10.

Can these both be true? The answer is a resounding and unconditional yes on the standard account of conditionals, which identifies indicative conditionals with material conditionals. It is not the same

---

I wish to thank Ranpal Dosanjh, Adam Harmer, Richard Gill, Adam Murray, Graham Priest, and two anonymous referees of The Journal of Philosophy for useful comments on earlier versions of this paper.

I say that indicative conditionals are contrary conditionals (in short, contraries), if they have the same antecedent but incompatible consequents. Likewise with counterfactual or material conditionals.
with many other contemporary accounts of conditionals. They yield a qualified negative answer: (a) and (b) are not compatible unless their common antecedent cannot be true. I think this is a wrong answer. The conditionals can both be true while their antecedent states a possibility. This is what I aim to show in this paper. In doing so, I present a solution to Raymond Smullyan’s intriguing version of the two envelope paradox.

1. Introduction

Consider a usual formulation of the two envelope paradox:

Suppose that there are two envelopes, Ali and Baba, and that you know that one of them has twice as much money as the other, but not which has more. If so, is it rational to choose one of them over the other? One might argue that Baba is preferable as follows. Suppose that Ali has $x$. Then Baba is as likely to have $2x$ as to have $x/2$. So its expected value is $1¼x$ ($¼ x + ½ x/2$), which is greater than Ali’s amount, $x$. So Baba is preferable. But the same reasoning leads to the opposite conclusion as well: Ali is preferable.

---

2 Two or more statements are compatible, if they can all be true (or they all hold on the same possible situation).


4 For convenience of exposition, I take this to imply that both have a positive amount.
This formulation of the paradox (or puzzle)\(^5\) involves assumptions about probability distribution, and invokes decision-theoretic principles that relate expected values (or utilities) to rationality of choice.\(^6\)

Smullyan formulates a version of the paradox that involves no such assumptions or principles. His version of the paradox is worthy of a special designation, because it differs significantly from the usual, probabilistic versions. Call it Smullyan’s paradox.

To formulate the paradox, Smullyan presents a possible situation (call it the Smullyan situation):

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. (189)\(^7\)

And he gives arguments for two incompatible theses on the situation: “Proposition 1. The amount

\(^5\)I use ‘paradox’ in a broad sense to include substantial puzzles as well as genuine paradoxes.


\(^7\)All the parenthetical references in this section are to Smullyan, op.cit. Note that the descriptions of the situation do not explicitly include the condition that the amounts in the envelopes are unknown. It is not necessary to assume this in formulating Smullyan’s paradox. See the arguments for Propositions 1 & 2 given below.
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” (190). Here is the 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 \( n/2 \). (190f)

And the 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. (191)

The Smullyan situation, which is not impossible, cannot satisfy both propositions. So at
least one of the arguments must be faulty. If so, which one is? I think both are. Both of them rest
on an incorrect assumption about the logic of conditionals.

To see this, note that both Proposition 1 and Proposition 2 involve two definite descriptions:

____________________

8I 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 the suggestion that a statement satisfied by
a situation (or possible world) must be in the situation (or possible world).
(α) the amount that you will gain if you gain;

(β) the amount you will lose if you lose.

These definite descriptions carry *uniqueness conditions*:

(a) If something is an *amount that you will gain if you gain* and something is an *amount that you will gain if you gain*, then the former amount is the same as the latter.

(b) If something is an *amount that you will lose if you lose* and something is an *amount that you will lose if you lose*, then the former amount is the same as the latter.

So both propositions imply (a) and (b). And we can see that the Smullyan situation must fail both propositions because it cannot satisfy both conditions.

To see this, suppose that Ali and Baba are the two envelopes in the situation with Ali being the one you hold. And suppose, for example, that Baba has more than Ali by $10 (so Ali has $10). Using these assumptions, we can derive two contrary conditionals:

(c) If you lose (on the trade), you will lose $5.

(d) If you lose (on the trade), you will lose $10.

Here are the derivations:

*Argument for (c)*: The envelope you hold, Ali, has $10. So if you lose on the trade, Baba
must have half as much, $5, and you will lose $5 (= $10 – $5).

*Argument for (d):* If you lose, you will lose the difference between the amounts in Ali and Baba. This is $10. So you will lose $10 if you lose.

And (c)–(d) imply that one of the uniqueness conditions, (b), is false. The same holds as long as Baba has more: assuming that Baba has more by a certain amount, we can derive cousins of (c) and (d) that imply the negation of (b). Similarly, (a) must fail if Ali has more.

So both of Smullyan’s arguments lead to incorrect conclusions. And we can see the mistakes made in the arguments to reach the conclusions. The first argument derives Proposition 1 from the two theses it reaches in the penultimate step:

\[
\begin{align*}
(e) & \quad \text{If you gain (on the trade), you will gain } n \text{ dollars.} \\
(f) & \quad \text{If you lose (on the trade), you will lose } n/2 \text{ dollars.}
\end{align*}
\]

To derive the proposition, the argument assumes that (e) and (f) yield identity statements with (α) and (β):\(^{10}\)

---

\(^9\)I often use en dash to indicate the so-called collective predication over the relevant term. So the above sentence, for example, means that (c) and (d), *taken together*, imply that (b) is false.

\(^{10}\)The brackets indicate that the condition clauses ‘if you gain’ and ‘if you lose’ take the narrow scope to yield identity statements involving the definite descriptions. Without the brackets, (e’) and (f’) have scope ambiguity. They cannot be considered identity statements if the clauses are taken to take the wide scope. The statement that results from regarding ‘if you gain’ in (e’) to take the wide scope, for example, is a conditional equivalent to ‘If you gain, the amount you will gain is \(n\) dollars.’ And this is not equivalent to (e’); unlike (e’), it does not imply (a).
(e')    [The amount that you will gain if you gain] is $n$ dollars.

(f')    [The amount that you will lose if you lose] is $n/2$ dollars.

But these do not follow from (e) and (f). Unlike (e) and (f), they imply (a) and (b). Moreover, one of these uniqueness conditions must fail, as we have seen. The argument for Proposition 2 makes the same error in the last step, where cousins of (e') and (f') are derived from those of (e) and (f).  

We can now see that Smullyan’s paradox relates to a significant issue about the logic of conditionals. On the above analysis of the paradox, (a), for example, can be false because two related contrary conditionals can both be true. Moreover, on the analysis, two contrary conditionals whose common antecedent states a possibility (e.g., (c) and (d)) can both be true. So the analysis rejects a thesis accepted by most major contemporary accounts of conditionals:

Exclusion Thesis: Contrary conditionals are incompatible if their antecedent states a possibility.

This is no surprise to those who accept the standard account of conditionals, which identifies indicative conditionals with material conditionals, for material conditionals are true as long as their antecedents are false. But the account is subject to serious challenges. And its major alternatives, 

---

11 For example, the relevant cousin of (e) is ‘If you gain, you will gain $d$ dollars’, and that of (e') is ‘[The amount you will gain if you gain] is $d$ dollars.’

12 The material conditional ($\phi \rightarrow \psi$) is true if, and only if, either $\phi$ is not true or $\psi$ is true.

13 For objections to the account, see Robert C. Stalnaker, “Indicative conditionals”, Philosophia 5, 3 (July 1975), reprinted in William L. Harper et al. (eds.), Ifs: Conditionals, Belief,
which take the probabilistic or the possible world approach,14 rest on ideas that directly lead to the exclusion thesis.\textsuperscript{15}

What does this mean? I think it means that the probabilistic and possible world approaches yield wrong accounts of conditionals. Proponents of the approaches might disagree. They might defend the exclusion thesis, and reject the solution to Smullyan’s paradox sketched above. To do so, they would have to offer plausible alternatives to the solution. I do not think they can. One cannot solve the paradox without rejecting the thesis. For Smullyan’s arguments are valid up to the penultimate steps.

To see this, it is necessary to draw a clear distinction between indicative and counterfactual conditionals. Smullyan uses indicative conditionals to formulate the paradox. And the penultimate conclusions of his arguments involve indicative conditionals (e.g., (e) and (f)). We can show that those conclusions follow from the descriptions of the Smullyan situation, and yield contrary indicative conditionals that form a counterexample to the Exclusion Thesis. It is not the same,


\textsuperscript{15}Strictly speaking, the probabilistic approach yields only a weaker thesis: contrary conditionals are incompatible if their antecedent states a possibility \textit{with a positive probability}. (Likewise with its indicative and counterfactual cousins stated below.) It is straightforward to see that the objection to the Exclusion Thesis and its indicative version given in this paper applies to their weaker cousins as well. So I ignore the difference for convenience of exposition. See note \textsuperscript{44}. 8
however, with counterfactuals. The descriptions of the situation, we shall see, do not imply the counterfactual cousins of the penultimate conclusions of Smullyan’s arguments. And counterfactual cousins of Smullyan’s paradox do not help to obtain counterfactual counterexamples to the thesis.

So it is useful to separate the exclusion thesis into two parts:

**Indicative Exclusion Thesis:** Contrary indicative conditionals are incompatible if their antecedent states a possibility.

**Counterfactual Exclusion Thesis:** Contrary counterfactual conditionals are incompatible if their antecedent states a possibility.

A proper solution of Smullyan’s paradox requires rejecting Indicative Exclusion, but not rejecting Counterfactual Exclusion. And we can give a convincing case for this thesis.

I present a solution to Smullyan’s paradox in the next section, and discuss its counterfactual cousins in sections 3 and 4. In section 5, I use the solution to the paradox to argue against the Indicative Exclusion Thesis. In section 6, I give a direct argument against the thesis, relate its failure to substitutivity of identity, and explain why the thesis fails while its counterfactual cousin holds. Section 7 gives a concluding summary.

**2. Smullyan’s paradox**

To give a thorough analysis of Smullyan’s paradox, it is useful to formulate a streamlined version
that highlights its logical backbone by removing talk of *decision, gain, and loss*.

To do so, imagine a possible 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. Now, let $f(a)$ and $f(b)$ be the amounts in dollars in Ali and Baba, respectively. Then we can turn Smullyan’s arguments for Propositions 1 & 2 to arguments for two incompatible theses:

$$[P1] \text{[The amount that is the difference between } f(a) \text{ and } f(b) \text{ if Baba has more] is greater than [the amount that is the difference between } f(a) \text{ and } f(b) \text{ if Ali has more].}$$

$$[P2] \text{[The amount that is the difference between } f(a) \text{ and } f(b) \text{ if Baba has more] is the same as [the amount that is the difference between } f(a) \text{ and } f(b) \text{ if Ali has more].}$$

We can divide the resulting arguments for these theses into two parts. The first parts reach counterparts of the penultimate conclusions of Smullyan’s arguments for Propositions 1 & 2:

$$[T1] \text{There is a number } n \text{ and a smaller number } m \text{ such that if Baba has more the difference between } f(a) \text{ and } f(b) \text{ is } n, \text{ but that if Ali has more the difference is } m.$$  

$$[T2] \text{There is a number } l \text{ such that if Baba has more the difference between } f(a) \text{ and } f(b) \text{ is } l, \text{ and that if Ali has more the difference is also } l.$$  

One might argue for these as follows:

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

\textit{Argument 2}: Let $l$ be the difference between $f(a)$ and $f(b)$. Then if Baba has more the difference is $l$; and if Ali has more the difference is also $l$.

In the second parts, one might conclude [P1] from [T1], and [P2] from [T2].

This gives rise to a paradox. The descriptions of $S$ are clearly consistent, but the final conclusions of the arguments, [P1] and [P2], are not. (Call the paradox the \textit{Simplified Smullyan}.)

And we can derive Smullyan’s paradox from the Simplified Smullyan by adding to the descriptions of $S$ the following conditions:

(i) You will gain if and only if Baba has more than Ali.
(ii) You will lose if and only if Ali has more than Baba.
(iii) You will gain $n$ if, and only if, Baba has more than Ali and the difference between $f(a)$ and $f(b)$ is $n$.
(iv) You will lose $n$ if, and only if, Ali has more than Baba and the difference between $f(a)$ and $f(b)$ is $n$.

Given these conditions, [P1] and [P2] are equivalent to Propositions 1 and 2, respectively,\footnote{Propositions 1 and 2 result essentially from replacing the right sides of (i) and (ii) in [P1] and [P2] with their left sides.} and [T1] and [T2] to their counterparts in Smullyan’s arguments:
[T1'] There is a number $n$, and a number $m$ smaller than $n$ such that if you gain you will gain $n$, but that if you lose you will lose $m$.

[T2'] There is a number $l$ such that if you gain you will gain $l$, and that if you lose you will lose $l$.

So the derivations of [P1] and [P2] from the descriptions of $S$ would yield derivations of Propositions 1 & 2 from the descriptions together with (i)–(iv). This also gives rise to a paradox, Smullyan’s paradox, for the additional conditions are consistent with the descriptions of $S$.

The two paradoxes have the same solution. Both the arguments for Propositions 1 & 2 and those for [P1] and [P2] break down in their second parts. It is wrong to infer [P1] and [P2], for example, from [T1] and [T2]. These theses are consistent; they both follow from the descriptions of $S$. But they (taken together) contradict both [P1] and [P2].

To see this, it is useful to formulate relevant statements in a regimented language. Let ‘$k$’, ‘$l$’, ‘$m$’, and ‘$n$’ be restricted variables for numbers; ‘$x$’, ‘$y$’, and ‘$z$’ restricted variables for envelopes on the table; and ‘$a$’ and ‘$b$’ singular constants for Ali and Baba, respectively. And let ‘$C$’ be a dyadic predicate whose meaning is given as follows:

$$xCn: x \text{ contains exactly } n.$$ 

Finally, let ‘$f(x)$’, ‘$d(x, y)$’, and ‘$x>y$’ abbreviate ‘$(\exists n)xCn^{17}$’, ‘$|f(x) – f(y)|$’, 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

\footnote{I use the lower-case inverted iota ‘$\gamma$’ as the definite description operator.}
dollar amounts in $x$ and $y$, respectively; and `$x > y$’ means: $x$ has more money than $y$.

We can now formulate the descriptions of $S$ as follows:

\[(0) \quad \exists x \exists y [x \neq y \land \forall z (z = x \lor z = y) \land f(x) = 2f(y) \land f(y) > 0].^{18} \] (There are exactly two envelopes on the table, and one of them has twice as much money as the other.)

\[(1) \quad \exists x \ x = a \land \exists x \ x = b \land a = b. \] (Ali and Baba are different envelopes on the table.)

And [T1] and [T2] can be formulated as follows: \(^{19}\)

\[\text{[T1]} \quad \exists n \exists m < n ([b > a \rightarrow d(a, b) = n] \land [a > b \rightarrow d(a, b) = m]). \] (There is a number $n$ and a smaller number $m$ such that if Baba has more the difference between $f(a)$ and $f(b)$ is $n$, but that if Ali has more the difference is $m$.)

\[\text{[T2]} \quad \exists l ([b > a \rightarrow d(a, b) = l] \land [a > b \rightarrow d(a, b) = l]). \] (There is a number $l$ such that if Baba has more the difference between $f(a)$ and $f(b)$ is $l$, and that if Ali has more the difference is also $l$.)

To formulate [P1] and [P2] in a regimented language, we may use two definite descriptions:

\[\text{(a)} \quad (\exists k) [b > a \rightarrow d(a, b) = k] \] (the amount that is the difference between $f(a)$ and $f(b)$ if...

---

\(^{18}\) Note that the conjunct ‘$x \neq y$’ is redundant (it is implied by ‘$f(x) = 2f(y) \land f(y) > 0$’).

\(^{19}\) I use ‘$\rightarrow$’ for the indicative conditional. I do not assume the account of the indicative conditional as the material conditional, for which I use ‘$\Rightarrow$’. See note \(23\).
Baba has more)

\( (\beta) \ (k)[a > b \rightarrow d(a, b) = k] \) (the amount that is the difference between \( f(a) \) and \( f(b) \) if Ali

has more)

We can then formulate the theses as follows:

\[
[P1] \ (k)[b > a \rightarrow d(a, b) = k] > (k)[a > b \rightarrow d(a, b) = k]). \\
[P2] \ (k)[b > a \rightarrow d(a, b) = k] = (k)[a > b \rightarrow d(a, b) = k]).
\]

Now, we can show that \([T1]\) and \([T2]\), taken together, contradict both \([P1]\) and \([P2]\). The
definite descriptions \((\alpha)\) and \((\beta)\) carry **uniqueness conditions**:\(^{20}\)

\[
[U1] \ \forall l\forall m([b > a \rightarrow d(a, b) = l] \land [b > a \rightarrow d(a, b) = m] \rightarrow l = m). \\
[U2] \ \forall l\forall m([a > b \rightarrow d(a, b) = l] \land [a > b \rightarrow d(a, b) = m] \rightarrow l = m).
\]

So both \([P1]\) and \([P2]\), which have both \((\alpha)\) and \((\beta)\), imply the conjunction of \([U1]\) and \([U2]\). But

\(^{20}\) The antecedents in \([U1]\) and \([U2]\) symbolize the following:

If Baba has more, the difference between \( f(a) \) and \( f(b) \) is \( l \); and if Baba has more, the
difference between \( f(a) \) and \( f(b) \) is \( m \).

If Ali has more, the difference between \( f(a) \) and \( f(b) \) is \( l \); and if Ali has more, the difference
between \( f(a) \) and \( f(b) \) is \( m \).
its negation follows from [T1]–[T2]. So these (taken together) contradict both [P1] and [P2].

This means that neither [T1] nor [T2] implies [P1] or [P2] unless [T1]–[T2] are inconsistent. And these are not inconsistent. Both of them follow from (0)–(1), which are clearly consistent.

To see this, note that (0)–(1) imply all the following:

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

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

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

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

And (1)–(4) imply [T1] and [T2]. We can show this by elaborating on the informal arguments presented above, Arguments 1 & 2, which draw parallels to the first parts of Smullyan’s arguments.

\[ \text{---------} \]

\[ ^{21}\text{Otherwise the following instances of [T1] & [T2] and [U1] & [U2] must all be jointly consistent:} \]

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

\[ [b \succ a \rightarrow d(a, b) = l] \land [a \succ b \rightarrow d(a, b) = l]. \]

\[ [b \succ a \rightarrow d(a, b) = l] \land [b \succ a \rightarrow d(a, b) = n] \rightarrow l = n. \]

\[ [a \succ b \rightarrow d(a, b) = m] \land [a \succ b \rightarrow d(a, b) = m] \rightarrow l = m. \]

But these imply a contradiction: ‘$m = n$’ and ‘$m < n$’.

\[ ^{22}\text{Incidentally, (1)–(4), which imply (0), are equivalent to (0)–(1).} \]
for Propositions 1 & 2.\textsuperscript{23} [T1] follows from (1), (3a), and (4):\textsuperscript{24}

\textit{Argument 1}: Let a positive number \( n \) be \( f(a) \) (there is such a number by (3a)). Then if \( b \succ a \),

\[ f(b) = 2n \quad \text{(by (1) & (4))} \text{ and } d(a, b) = f(b) - f(a) = 2n - n = n. \] And if \( a \nrightarrow b \), \( f(b) = n/2 \) (by (1) & (4)) and \( d(a, b) = f(a) - f(b) = n - n/2 = n/2. \) And \( n > n/2 \).

And [T2] follows from (3a)–(3b):

\textit{Argument 2}: Let \( n \) be \( |f(a) - f(b)| \) (there is such a number by (3a) & (3b)\textsuperscript{25}). Then if \( b \succ a \),

\[ d(a, b) = f(b) - f(a) \text{ and } f(b) - f(a) = |f(b) - f(a)| = n. \] And if these hold, \( d(a, b) = n. \)

\textsuperscript{23}The arguments given below are somewhat more involved than Arguments 1 & 2; they are meant to show that it is not necessary to use controversial rules of inference that result from identifying indicative conditionals with material conditionals, such as: (a) \( \prec \phi \quad \phi \nrightarrow \psi \); and (b) \( \psi - \phi \nrightarrow \psi \). Except for substitutivity of identity, the arguments use the following rules:

- (i) \( \phi \cdot \psi, \psi \cdot \chi + \phi \cdot \chi \).  
- (ii) \( \phi \cdot \psi + \phi \cdot (\psi \nrightarrow \chi) \).  
- (iii) \( \phi \cdot \psi, \phi \cdot \chi + \phi \cdot (\psi \nrightarrow \chi) \).

Stalnaker rejects (i) (see op. cit., p. 48), but his counterexamples concern only its counterfactual cousin, which I agree is not sound. Instead of (i), in any case, one can use a variant:

- (i') \( \phi \cdot (\phi \nrightarrow \psi), (\psi \nrightarrow \phi) \cdot \chi + \phi \cdot \chi \).

Note that the counterfactual cousin of (i') is also sound.

\textsuperscript{24}Similarly, (1), (3b), and (4) imply the mirror image of [T1]:

[T3] \( \exists m \in \mathbb{N} \forall n \in \mathbb{N} [b \nrightarrow a \rightarrow d(a, b) = m] \land [a \nrightarrow b \rightarrow d(a, b) = n] \).

\textsuperscript{25}(3a) and (3b) imply that ‘\( f(a) \)’ and ‘\( f(b) \)’ (which abbreviate definite descriptions) are legitimate terms referring to definite amounts (or numbers).
Similarly, if $a \succ b$, $d(a, b) = f(a) - f(b)$ and $f(a) - f(b) = |f(a) - f(b)| = n$. And if these hold, $d(a, b) = n$.

$(0)–(1)$ imply $[T1]$ and $[T2]$, we have seen, while these (taken together) imply the negations of both $[P1]$ and $[P2]$. And $(0)–(1)$ are consistent. So $[T1]–[T2]$ imply neither $[P1]$ nor $[P2]$. And both of these must fail on the possible situation $S$, which satisfies $(0)$ and $(1)$.

This solves the Simplified Smulynan. Smulynan’s paradox has the same solution. Given the additional conditions (i)–(iv), $[T1']$ and $[T2']$ follow from $(0)–(1)$; they, given the conditions, are equivalent to $[T1]$ and $[T2]$, respectively. And $[T1']–[T2']$ contradict Propositions 1 & 2; both propositions imply the uniqueness conditions for the definite descriptions ‘the amount you will gain if you gain’ and ‘the amount you will lose if you lose’, but the conjunction of the conditions conflicts with $[T1']–[T2']$. So these theses imply neither proposition. And both propositions must fail on the Smulynan situation, which satisfies both $[T1']$ and $[T2']$.

3. Indicatives or Counterfactuals?

The above solution to Smulynan’s paradox takes it to be formulated with indicative conditionals. One might attempt to formulate a related paradox that defies the solution by reformulating the paradox with counterfactual conditionals. I discuss such attempts and give solutions to the resulting paradoxes in the next section. In this section, I prepare for the discussion by examining an analysis of Smulynan’s paradox that takes it to involve counterfactuals.
Chase disputes Proposition 1 in his discussion of Smullyan’s paradox. He argues that it cannot hold on the Smullyan situation. To do so, however, he does not distinguish the proposition with the intermediate thesis [T1’], and formulates it as follows:

(C) There are \(x\) and \(y\) such that . . . if you gain, you gain \(x\) . . . if you lose, you lose \(y\), and . . . \(x > y\). (158)

Thus, he in effect challenges the first part of Smullyan’s argument for the proposition. To do so, he holds that the two conditionals in (C) 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” (158). Then he argues that (C), or its counterfactual cousin, must fail on the Smullyan situation.

Chase is right to note that the antecedents of the conditionals in (C) are incompatible. But this does not mean that “there is no way they can both be . . . interpreted” as indicatives (160); nor does it mean that the statement, if taken to have indicatives, must be false or fail to hold on the Smullyan situation. Consider a statement with the same structure:

There are integers \(m\) and \(n\) such that \(\chi(\pi) = m\) if \(\pi\) is a rational number, \(\chi(\pi) = n\) if \(\pi\) is not a rational number, and \(m > n\),

\(^{26}\)See James Chase, “The non-probabilistic two envelope paradox”, Analysis 62, 2 (April 2002): 157-60. All the parenthetical references in this section are to this article.
where \( \chi_{r} \) is for the characteristic function for the property of being a rational number. Surely, both conditionals in the statement can be taken to be indicatives although their antecedents contradict each other. Moreover, the statement is true when they are so taken; it follows from ‘Any real number \( r \) is such that \( \chi_{r}(r) = 1 \) if \( r \) is a rational number, and that \( \chi_{r}(r) = 0 \) if \( r \) is not a rational number.’ If so, there is no reason not to take the conditionals in (C) to be indicatives. And the statement as so understood (i.e., \([T1']\)) must hold on the Smullyan situation or any possible situation that satisfies the descriptions thereof; (0)–(1) together with (i)–(iv) imply \([T1']\) as (0)–(1) imply \([T1]\).

Is it the same with the counterfactual cousin of (C) or \([T1']\)? Does the statement (call it \('[C*]'\)) also follow from the descriptions of the Smullyan situation? I think not.

To see this, consider a situation that extends the Smullyan situation by satisfying additional conditions:

Ali’s amount and Baba’s amount were fixed as $10 and $20, respectively, before you made a choice between them. And you chose Ali although you were as likely to choose Baba as to choose Ali.

The situation (or a possible world that extends it) satisfies all the descriptions of the Smullyan situation, but not \([C*]\). It does not satisfy the conjunction of any matching instances of the two counterfactuals in \([C*]\), such as the following:

\[27\text{For more on indicative conditionals with false antecedents, see Graham Priest, “Rational dilemmas”, Analysis 62, 1 (January 2002): 11-16 (p. 14f).}\

\[28\text{That is, ‘There is a number } n \text{, and a number } m \text{ smaller than } n \text{ such that if you were to gain you would gain } \$n \text{, but that if you were to lose you would lose } \$m.’}\]
(a) If you were to gain (on the trade), you would gain $10.

(b) If you were to lose (on the trade), you would lose $5.

Although it satisfies (a), it does not satisfy (b): if you were to lose, it would be because you had chosen Baba and thus you would lose $10, not $5. So the descriptions of the Smullyan situation fail to imply \([C^*]\).

If so, do the descriptions of the Smullyan situation contradict \([C^*]\) as they contradict Proposition 1? According to Chase, the answer is yes. He argues, as noted above, that \([C^*]\) must fail on the Smullyan situation: it cannot hold on any possible world extending the situation, for “only one of [the conditionals in it] can be true [on the world]” (159). I disagree.

\([C^*]\) hold on some situations that satisfy (0), (1), and (i)–(iv). Let Smullyan* be such a situation in which $10 was put in Ali, which was given to you, and then the amount in Baba was determined to be $20 or $5 by a toss of a coin. This situation satisfies \([C^*]\), for it satisfies both (a) and (b): if you were to gain, it would be because Baba had $20; if you were to lose, it would be because Baba had $5.

Chase might object that Smullyan* is not an extension of the Smullyan situation, where you pick up Ali after the amounts in the envelopes have been fixed. He might argue that this makes a difference: no situation that extends the Smullyan situation can satisfy \([C^*]\). Suppose that (a), for example, holds (on such a situation) because Ali and Baba have $10 and $20, respectively (in the situation). Then (b) must fail, for if you were to lose on the trade, you would have chosen Baba instead of Ali (while their amounts remain the same), and thus lose $10, not $5.\(^{29}\) But we can specify

\(^{29}\)See Chase, *op.cit.*, p. 159 for his argument that \([C^*]\) must fail on the Smullyan situation.
an extension of the Smullyan situation that satisfies [C*] as well. Consider the following conditions:

First, Ali’s amount was fixed as $10. Then Baba’s was determined to be $5 or $20 by a toss of a coin. And then you made a choice between the envelopes, but you were pretty much pre-determined to choose Ali. (For example, you have a strong, perhaps unknown, inclination for envelopes of a certain kind that makes you choose Ali.)

An extension of the Smullyan situation that satisfies these conditions satisfies [C*] as well. For it satisfies both (a) and (b): if you were to gain on the trade, Baba would have had $20; but if you were to lose on the trade, Baba would have had $5.

We have seen that the descriptions of the Smullyan situation neither implies nor contradicts [C*], the counterfactual cousin of [T1']. Can we turn Smullyan’s argument for [T1'] to an argument for [C*] by using counterfactuals instead of indicatives in describing the situation? If so, one might attempt to obtain a cousin of Smullyan’s paradox that defies the solution presented in the last section. I examine attempts to do so in the next section.

4. Counterfactual Cousins of Smullyan’s Paradox

Consider the counterfactual cousins of [T1] and [T2]:

\[30\]

\[30\]I use ‘\(\neg a \rightarrow \neg b\)’ for the counterfactual conditional, usually indicated in English by the subjunctive mood. \([a \neg b \rightarrow d(a, b) = n]\), for example, amounts to ‘If Ali had more than Baba, the difference between the amounts in Ali and Baba would have been \(n\).’
\[T1^*\] \(\exists n \exists m < n([b > a \rightarrow d(a, b) = n] \land [a > b \rightarrow d(a, b) = m])\). (There is a number \(n\) and a smaller number \(m\) such that if Baba had more the difference between \(f(a)\) and \(f(b)\) would have been \(n\), but that if Ali had more the difference would have been \(m\).)

\[T2^*\] \(\exists l([b > a \rightarrow d(a, b) = l] \land [a > b \rightarrow d(a, b) = l])\). (There is a number \(l\) such that if Baba had more the difference between \(f(a)\) and \(f(b)\) would have been \(l\), and that if Ali had more the difference would also have been \(l\).)

Can we obtain a counterfactual cousin of Smullyan’s paradox by turning Arguments 1’ & 2’ to arguments for these theses?

Note that the counterfactual cousins of the arguments are not valid. They have two major problems, which stem from logical differences between indicatives and counterfactuals.

First, (0)–(1) do not imply the counterfactual cousins of most indicative conditionals they imply. Consider, for example, the following pair:

(a) \([a > b \rightarrow f(a) = 2f(b)]\). (If Ali has more, Ali has twice as much as Baba.)

(a*) \([a > b \rightarrow f(a) = 2f(b)]\). (If Ali had more, Ali would have had twice as much as Baba.)

(0)–(1) imply the indicative (a), but not the counterfactual (a*). But the counterfactual cousin of Argument 1’ needs (a*) because Argument 1’ uses (a).

Second, substitutivity of identity holds in the indicative context, but not in the counterfactual context.

---

\(^{31}\)(0) implies (4), ‘\(\forall x \forall y [x > y \rightarrow f(x) = 2f(y)]\).’ And this, given (1), implies (a).
context. Suppose that *Ali’s amount* is $20, and that *Ali’s amount* would have been less than $10 if Ali had less than Baba. And yet one cannot conclude that $20 would have been less than $10 if Ali had less than Baba.32 This gives rise to a problem for the counterfactual cousins of Arguments 1’ & 2’, which invoke substitutivity of identity for indicative conditionals. For example, Argument 1’ involves using the identity ‘f(a) = n’ to substitute ‘n’ for ‘f(a)’ in the indicative (a). But one cannot invoke the identity to make the same substitution in the counterfactual (a*).

We can avoid these problems by adding some conditions to (0)–(1) or by replacing them with stronger conditions. Consider the following conditions:

*Counterfactual Smullyan Conditions:*

(0*)  \[ \exists x \exists y [x \neq y \land \forall z (z = x \lor z = y) \land \square (f(x) = 2 \cdot f(y) \land f(y) > 0)] \]  (There are exactly two envelopes on the table, and one of them must have twice as much as the other.)

(1)  \[ \exists x : x = a \land \exists x : x = b \and a \neq b \]  (Ali and Baba are different envelopes on the table.)

(5)  \[ \exists n \square f(a) = n \]  (Ali must have a fixed amount.)

(6)  \[ \exists n \square d(a, b) = n \]  (Ali’s amount and Baba’s must differ by a fixed amount.)

These conditions imply not only the indicative (a) but also the counterfactual (a*); (0*), which is stronger than (0), implies the counterfactual cousin of (4):

32By contrast, the indicative cousin of the inference is unexceptionable: supposing that *Ali’s amount* is $20, and that *Ali’s amount* is less than $10 if Ali’s amount is less than Baba’s (for this is $10), one can conclude that $20 is less than $10 if Ali’s amount is less than Baba’s.
(4*) $\forall x \forall y [x \rightarrow y \rightarrow f(x) = 2 \cdot f(y)]$. (If an envelope on the table were to have more than another, the former would have had twice as much as the latter.)

And we can use the necessary identity (5) to substitute ‘n’ for ‘f(a)’ in counterfactuals (e.g., (a*)).

So we can show that the above conditions imply both [T1*] and [T2*] by modifying Arguments 1' & 2'.

Now, some might infer from [T1*] and [T2*] the counterfactual cousins of [P1] and [P2]:

\[ [P1*] \quad (\exists)[b \rightarrow a \rightarrow d(a, b) = k] > (\exists)[a \rightarrow b \rightarrow d(a, b) = k]. \]

\[ [P2*] \quad (\exists)[b \rightarrow a \rightarrow d(a, b) = k] = (\exists)[a \rightarrow b \rightarrow d(a, b) = k]. \]

This would give rise to a paradox, because [P1*]–[P2*] are inconsistent while the Counterfactual Smullyan Conditions are consistent.

This paradox has the same solution as Smullyan’s. Just as [T1]–[T2] imply neither [P1] nor [P2], so do [T1*]–[T2*] imply neither [P1*] nor [P2*]. And for the same reason. Both [P1*] and

\[ \quad \text{-----------------------------} \]

33To see this, it is necessary to use counterfactual cousins of the rules of inference used in Arguments 1' & 2' (see note 23):

(i*) $\phi \rightarrow \neg (\phi \land \psi), (\phi \land \psi) \rightarrow (\phi \rightarrow \psi) \rightarrow \chi$. 

(ii*) $\phi \rightarrow \psi \rightarrow (\phi \rightarrow \psi)$.

(iii*) $\phi \rightarrow \neg \psi, \phi \rightarrow \neg \chi \rightarrow (\psi \lor \chi)$.

Note that (i*) is the counterfactual counterpart of (i'), not of (i).

34The two terms in the statements symbolize ‘the amount that would have been the difference between $f(a)$ and $f(b)$ if Baba had more’ and ‘the amount that would have been the difference between $f(a)$ and $f(b)$ if Ali had more’.  

24
[P2*] imply the uniqueness conditions for the two definite descriptions figuring in them.\textsuperscript{35}

\begin{align*}
[U1^*] & \forall l \forall m ([b \rightarrow a \Box \rightarrow d(a, b) = l] \land [b \rightarrow a \Box \rightarrow d(a, b) = m] \rightarrow l = m). \\
[U2^*] & \forall l \forall m ([a \rightarrow b \Box \rightarrow d(a, b) = l] \land [a \rightarrow b \Box \rightarrow d(a, b) = m] \rightarrow l = m). 
\end{align*}

But the conjunction of these is incompatible with [T1*]–[T2*].\textsuperscript{36} So these (taken together) imply negations of both [P1*] and [P2*]. And because [T1*]–[T2*] are consistent (they follow from the Counterfactual Smulian Conditions), they imply neither [P1*] nor [P2*].

So the attempt to formulate a paradox by deriving [P1*] and [P2*] from a consistent class of conditions via [T1*] and [T2*] is self-refuting. To succeed in deriving both [T1*] and [T2*] from such conditions is to show that they imply neither [P1*] nor [P2*].

Some might object that one can derive [P1*] and [P2*] from [T1*] and [T2*] by making a modest assumption:

\textsuperscript{35}The antecedents in [U1*] and [U2*] symbolize the following:

If Baba had more, the difference between \( f(a) \) and \( f(b) \) would have been \( l \); and if Baba had more, the difference between \( f(a) \) and \( f(b) \) would have been \( m \).

If Ali had more, the difference between \( f(a) \) and \( f(b) \) would have been \( l \); and if Ali had more, the difference between \( f(a) \) and \( f(b) \) would have been \( m \).

Note that the last conditional signs in [U1*] and [U2*], unlike the others, are indicative.

\textsuperscript{36}Given [U1*]–[U2*], the following instances of [T1*] and [T2*] are incompatible:

\begin{align*}
[b \rightarrow a \Box \rightarrow d(a, b) = n] \land [a \rightarrow b \Box \rightarrow d(a, b) = m] \land m < n. \\
[b \rightarrow a \Box \rightarrow d(a, b) = l] \land [a \rightarrow b \Box \rightarrow d(a, b) = l]. 
\end{align*}
(7) \(\Diamond a \rightarrow b \land \Diamond b \rightarrow a\). (Ali might have more than Baba, and Baba more than Ali.)

To do so, they might use a plausible thesis about counterfactuals:

**Counterfactual Exclusion Thesis ([CE]):** Contrary counterfactuals are incompatible if their antecedent states a possibility (in symbols, \(\Diamond \phi, \phi \Box \rightarrow \psi, \phi \Box \rightarrow \chi = \Diamond (\psi \land \chi)^{37}\)).

Invoking this thesis, which I think is correct (see section 6), one can derive the uniqueness conditions, [U1*] and [U2*], from (7).^{38} Given the conditions, [T1*] and [T2*] imply [P1*] and [P2*], respectively.\(^{39}\)

If so, can we devise a paradox by adding (7) to the Counterfactual Smullyan Conditions? The conditions (taken together) imply [P1*] and [P2*] (they imply (7) as well as [T1*] and [T2*]), but [P1*]–[P2*] are inconsistent. But this paradox has a simple solution. The Counterfactual Smullyan

---

^{37}This states that \(\Diamond \phi, \phi \Box \rightarrow \psi, \) and \(\phi \Box \rightarrow \chi\) imply \(\Diamond (\psi \land \chi)\), and is equivalent to the following:

\[\Diamond \phi, (\phi \Box \rightarrow \chi), \neg (\Diamond (\psi \land \chi)) = \neg (\phi \Box \rightarrow \psi)\].

^{38}By [CE], '\(\Diamond b \rightarrow a\)' and the antecedent of [U1*] imply '\(\Diamond [d(a, b) = l \land d(a, b) = m]\)', which implies '\(l=m\).' So '\(\Diamond b \rightarrow a\)' implies [U1*] if [CE] holds. Similarly, '\(\Diamond a \rightarrow b\)' implies [U2*] if [CE] holds.

^{39}Given the uniqueness conditions, the following, for example, are equivalent:

(i) \(a \rightarrow b \Box \rightarrow d(a, b) = n\).
(ii) \(\langle k \rangle[a \rightarrow b \Box \rightarrow d(a, b) = k] = n\).

For (ii) is equivalent to '\([a \rightarrow b \Box \rightarrow d(a, b) = n] \land \forall k([b \rightarrow a \Box \rightarrow d(a, b) = k] \rightarrow k=n)\).'
Conditions contradict (7), because they imply the following:\(^{40}\)

\[(8) \quad \Box a > b \lor \Box b > a. \text{ (Ali must have more than Baba or Baba must have more than Ali.)}\]

Moreover, \([T1^*]–[T2^*]\) imply (8) if Counterfactual Exclusion holds. So any conditions that imply (7) as well as \([T1^*]\) and \([T2^*]\) must be inconsistent if the thesis holds.

It is notable that although (7) is not consistent with the Counterfactual Smullyan Conditions, it is consistent with (0)–(1). This invites two questions about what happens if we add (7) to (0)–(1):

We have seen that (0)–(1) imply \([T1]\) and \([T2]\), and we can show that these together with (7) imply \([P1]\) and \([P2]\) by invoking the Indicative Exclusion Thesis (i.e., the indicative cousin of Counterfactual Exclusion). But \([P1]–[P2]\) are not consistent while (0), (1), and (7) are. If so, what is wrong with the derivations of \([P1]\) and \([P2]\) from these conditions?

Suppose that a possible situation satisfies (7) as well as (0) and (1). Then it must fail \([T1^*]\) or \([T2^*]\) (if Counterfactual Exclusion holds). Which of the two would it fail?

The answer to the first question is that the Indicative Exclusion Thesis is false. This is discussed in the next section. The rest of this section is devoted to answering the second question.

\(^{40}\)For (0*), (1), and (5) imply the following:

\[\exists n > 0(\Box [f(a) = n \land f(b) = 2n] \lor \Box [f(a) = n \land f(b) = n/2]). \text{ (There is a positive number } n \text{ such that either } f(a) \text{ must be } n \text{ while } f(b) \text{ is } 2n \text{ or } f(a) \text{ must be } n \text{ while } f(b) \text{ is } n/2.)\]
The answer to the question is that one cannot say which of [T1*] and [T2*] fails without more information about the situation. There are various kinds of possible situations that satisfy (0), (1), and (7). Some of them fail [T1*], others [T2*], and yet others both. It would be useful to see representative examples of the various kinds of situations.

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

$S_1$: Ali’s amount was fixed as $10, but Baba’s 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 (i) $10 for Ali and $20 for Baba or (ii) $20 for Ali and $10 for Baba. The choice between the two combinations was made by one toss of a fair coin: the first combination for heads, the second for tails. (The coin fell heads.)

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

Then $S_1$ satisfies [T1*] but not [T2*] (it satisfies (0*) and (5) but fails (6)). By contrast, $S_2$ satisfies

Also assume that the situations do not have peculiar features (e.g., the results of coin tossing having been fixed beforehand).
[T2*] but not [T1*] (it satisfies (6*) but fails both (0*) and (5)). Finally, $S_1$ fails both [T1*] and [T2*]. \(^{42}\) It is a situation in which Ali might have had $40–so the difference between their amounts might not have been $5 or $10, but $20 or more. \(^{43}\)

### 5. The Indicative Exclusion Thesis

Can contrary conditionals be compatible? One cannot hold that any contrary conditionals must be incompatible. Both $[(\phi \land \neg \phi) \rightarrow \phi]$ and $[(\phi \land \neg \phi) \rightarrow \neg \phi]$, for example, are true. But some might hold that any contrary conditionals whose common antecedents state a possibility must be incompatible. Call such contrary conditionals *strongly contrary conditionals* (in short, *strong contraries*). For example, ‘If you gain, you will gain $10’ and ‘If you gain, you will gain $5’ are strong contraries if it is possible for you to gain. Then we can formulate the thesis in question as follows:

\[
\text{Exclusion Thesis: Strongly contrary conditionals are incompatible.}
\]

This thesis also has an apparent problem. It clearly does not hold for *material* conditionals, which are true as long as their antecedents are false. But some might restrict the thesis to conditionals

\(^{42}\)The mirror image of $S_1$ (where Baba’s amount is fixed) is also a situation that satisfies neither [T1*] nor [T2*], but it satisfies the mirror image of [T1*]:

\[
\exists n \forall m < n ([b \rightarrow a \land d(a, b) = m] \land [a \rightarrow b \land d(a, b) = n]).
\]

$S_1$ violates this condition as well.

\(^{43}\)Given Counterfactual Exclusion, [P1*] and [P2*] are equivalent to [T1*] and [T2*], respectively. So $S_1$ satisfies only [P1*], $S_2$ only [P2*], and $S_3$ neither [P1*] nor [P2*].
commonly used in natural languages while pointing out that it is controversial whether any such conditionals (e.g., the indicatives) can be identified with material conditionals. They would then hold two theses that result from restricting the thesis to indicatives and counterfactuals:

**Indicative Exclusion Thesis ([IE]):** Strongly contrary indicative conditionals are incompatible (in symbols, $\diamond \phi, \phi \cdot \psi, \phi \rightarrow \chi = \diamond (\psi \land \chi)$).

**Counterfactual Exclusion Thesis ([CE]):** Strongly contrary counterfactual conditionals are incompatible (in symbols, $\diamond \phi, \phi \Box \rightarrow \psi, \phi \Box \rightarrow \chi = \diamond (\psi \land \chi)$).

These theses do not have an equal standing in contemporary debates about conditionals. Although the Counterfactual Exclusion Thesis has not been subject to serious challenge, the Indicative Exclusion Thesis has been. The main reason for this is that there are competing and divergent approaches to the analysis of indicative conditionals. The Indicative Exclusion Thesis is clearly wrong on the standard account of indicatives, which identifies them with material conditionals. Contrary indicatives, on the account, are equally true if their antecedent is false. But this is denied by major alternatives to the standard account. They take the probabilistic or the possible world approach, and the approaches rest on ideas that directly yield the thesis.\(^{44}\)

\(^{44}\)The probabilistic approach yields only a weaker thesis: Strongly contrary conditionals are incompatible if their antecedent has a positive probability. (A statement that states a possibility may not have a positive probability.) But the objection to the Indicative Exclusion given below applies *mutatis mutandis* to the indicative version of the weaker thesis:

**Weaker Indicative Exclusion:** Strongly contrary conditionals are incompatible if their antecedent has a positive probability (in symbols, $\diamond \phi$, Prob($\phi$)$>0$, $\phi \cdot \psi, \phi \rightarrow \chi = \diamond (\psi \land \chi)$).
To present the idea of the probabilistic approach, Ramsey suggests:

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 $q$.  

Stalnaker reformulates this idea for the possible world approach:

Consider a possible world in which $A$ is true, and which otherwise differs minimally from the actual world. “If $A$, then $B$” is true . . . just in case $B$ is true . . . in that possible world.

Strongly contrary conditionals are incompatible on this idea; similarly, one cannot simultaneously accept two such conditionals on Ramsey’s idea. So both Ramsey and Stalnaker hold versions of the Exclusion Thesis. Ramsey holds that “in a sense ‘If $p, q$’ and ‘If $p, \overline{q}$’ are contradictories” as long as the antecedent has a positive probability. And Stalnaker argues that “the denial of a conditional is equivalent to a conditional with the same antecedent and opposite consequent (provided that the antecedent is not impossible)” to explain an axiom in his system of conditional logic. Although

For we may assume, plausibly, that the antecedents of the conditionals in question state not only a possibility but one with a positive probability.


$^{46}$Stalnaker, *op.cit.*, p. 45 (original italics).

$^{47}$Ramsey, *op.cit.*, p. 247. (He uses the overline as the negation sign.)

$^{48}$Stalnaker, *op.cit.*, p. 48f. (See his Axiom (a4).) The informal statement is stronger than the formal axiom. Harper calls the former “Stalnaker’s axiom”; see “A sketch of some recent
these statements assert exclusion only between a special kind of strong contraries (i.e., those whose consequences are negations of each other), they would take the statements to imply exclusion between other kinds of strong contraries (e.g., ‘If you gain, you will gain $10’ and ‘If you gain, you will gain $5’) as well.\(^{49}\)

Now, Ramsey and Stalnaker do not distinguish indicatives from counterfactuals, and apply their approaches to both.\(^{50}\) So they hold Indicative Exclusion as well as Counterfactual Exclusion. I think that they are wrong to hold the former, although right to hold the latter.

We can see that Indicative Exclusion is false by applying the analysis of Smullyan’s paradox presented above. The falsity of the thesis follows from the following:

Thesis \(A\). (0), (1), and (7) imply both \([P1]\) and \([P2]\), if Indicative Exclusion holds.

Thesis \(B\). (0), (1), and (7) do not imply both \([P1]\) and \([P2]\).

Thesis \(B\) holds because (0), (1), and (7) are consistent while \([P1]\) and \([P2]\) are not. And we can use the analysis of Smullyan’s paradox to show that Thesis \(A\) holds. (0)–(1) imply \([T1]\) and \([T2]\), as the developments in the theory of conditionals”, in Harper et al. (eds.), \textit{op.cit.}, pp. 3-38 (p. 6). Although David Lewis rejects the counterfactual version of the informal statement, he accepts the counterfactual version of the formal axiom: \(\Diamond \phi, (\phi \square - \psi) = \sim (\phi \square - \sim \psi)\). See his \textit{Counterfactuals} (Cambridge, Harvard University Press, 1973), p. 16.

\(^{49}\)And their accounts yield the Exclusion Thesis (or its weaker, probabilistic sibling). It is straightforward to show this directly. One can also see it by relating the thesis to a version of the possibility preservation thesis discussed in the next section. It is the same with Lewis’s account of counterfactuals and \([CE]\). See note \(^{60}\).

\(^{50}\)Stalnaker makes this explicit in “Indicative conditionals”, rp. in Harper et al. (eds.), \textit{op.cit.}, pp. 193-210.
If \([IE]\) holds, \('b → a → d(a, b) = l \) \& \([b → a → d(a, b) = m] \to l = m\). \(\) and \('[b → a → d(a, b) = l] \& [b → a → d(a, b) = m] \to l = m\).  

Now, \([U2]\) implies the equivalence between the following:  

(a) \(a → b → d(a, b) = n\).

(a’) \((k)[a → b → d(a, b) = k] = n\).

For \((a’\) is equivalent to the conjunction of \((a) and \([U2]\). Similarly, \([U1]\) implies the equivalence between siblings of \((a) and \((a’):\)

(b) \(b → a → d(a, b) = n\)

(b’) \((k)[b → a → d(a, b) = k] = n\).

\[51\]If \([IE]\) holds, \('\diamond [b → a → d(a, b) = l] \& [b → a → d(a, b) = m] \to d(a, b) = m\)', and this implies \('\diamond [d(a, b) = l \& d(a, b) = m]\) which implies \('l = m\). So \((7)\) implies \([P1]\) if \([IE]\) holds. Likewise with \([P2]\).

\[52\]They symbolize ‘If Ali has more the difference between the amounts in Ali and Baba is $n’ and ‘[The amount that is the difference between the amounts in Ali and Baba if Ali has more] is $n.’
Using these equivalences, one can derive [P1] from [T1], and [P2] from [T2].

The analysis of Smullyan’s paradox, we have seen, yields a refutation of the Indicative Exclusion Thesis. And by applying the arguments given above for [P1] and [P2], we can see that any possible situation that satisfies (0), (1), and (7) is a counterexample to the thesis.\(^{53}\)

Suppose that Ali has $10 and Baba $20 in a possible situation that satisfies the three conditions. Then the situation satisfies ‘\(f(a) = 10\)’ and ‘\(f(b) = 20\)’. And these together with (0)–(1) imply strong contraries for the situation.\(^{54}\)

\[
\begin{align*}
(c) \quad a \not\rightarrow b & \rightarrow d(a, b) = 5. \\
(d) \quad a \not\rightarrow b & \rightarrow d(a, b) = 10.
\end{align*}
\]

We can see this by running mutatis mutandis the second parts of Arguments 1’ & 2’:

**Argument 1’**: Assume (0) and (1). And let \(f(a)\) be 10. Then if \(a \not\rightarrow b, f(b) = f(a)/2\) (by (1) & (4)) and \(d(a, b) = f(a) - f(b) = f(a)/2 = 5\). So if \(a \not\rightarrow b, d(a, b) = 5\).

**Argument 2’**: Let \(f(a) = 10\) and \(f(b) = 20\). Then \(|f(a) - f(b)| = 10\). And if \(a \not\rightarrow b, d(a, b) = f(a) - f(b) = |f(a) - f(b)| = 10\). If these hold, \(d(a, b) = 10\). So if \(a \not\rightarrow b, d(a, b) = 10\).

\(^{53}\)I assume that a possible situation for (0)–(1) must assign specific values as the amounts in Ali and Baba.

\(^{54}\)The contrary conditionals are strong contraries for the situation, because it is possible for their antecedent to hold on the situation (it satisfies (7)).
Similarly, the mirror image of the situation (where Ali has $20 and Baba $10) satisfies the mirror images of (c) and (d): ‘\( b \succ a \rightarrow d(a, b) = 5 \)’ and ‘\( b \succ a \rightarrow d(a, b) = 10 \).’ And it is the same with any possible situations that satisfy (0), (1), and (7). They must satisfy either strongly contrary substitution instances of (a) or such instances of (b).\(^{55}\)

This means that one can easily “create” or “realize” counterexamples to Indicative Exclusion, possible situations satisfying strongly contrary indicatives. For the conditions to jointly realize (i.e., (0), (1), and (7)) are quite weak.\(^{56}\) Moreover, we can see that realizing even weaker conditions would do. Let (0′) be a statement resulting from replacing the numeral ‘2’ in (0) with the numeral for any positive number except 1. Then essentially the same argument shows that any possible situation that satisfies (0′) as well as (1) and (7) satisfies strong indicative contraries. So simply pick two envelopes, put different amounts in them, and place them on a table with no other envelope. Then you have made two strong contraries true.

6. Indicatives and Counterfactuals

We cannot solve Smullyan’s paradox without rejecting Indicative Exclusion, as we have seen, but we can solve its counterfactual cousins without rejecting Counterfactual Exclusion. And I think this thesis is correct. What gives rise to this divergence between indicatives and counterfactuals?

To see the reason for the divergence, recall that substitutivity of identity is valid for

\(^{55}\)By substitution instances of (a) or (b), I mean sentences one can obtain from them by replacing the variable ‘\( n \)’ with numerals.

\(^{56}\)Note that no conditional is used in their formulations.
indicatives, though not for counterfactuals. Consider, for example, the following:

(a) \textit{Ali’s amount} is $10.

(b) If Ali has more than $20, \textit{Ali’s amount} is larger than $20.

(c) If Ali has more than $20, $10 is larger than $20.

(c) follows from (a)–(b). Although its consequent is false, this does not mean that it is wrong to use (a) to substitute ‘$10’ for ‘Ali’s amount’ in (b), but that the antecedent of (c) must also be false if both (a) and (b) are true. In drawing this conclusion, we assume that these imply (c).

We can now see the reason for the failure of Indicative Exclusion. Consider a possible situation, $S*$, in which Ali happens to have $10 although it might have had more than $20.\textsuperscript{57} (c) must hold on the situation; it satisfies both (a) and (b). And while its antecedent states a possibility for the situation, its consequent does not. So the situation is a counterexample to Indicative Exclusion. For the thesis implies that indicative conditionals preserve possibility.\textsuperscript{58}

\textit{Indicative Preservation of Possibility ([IPP])}: If the antecedent of a true indicative states a possibility, its consequent must also state a possibility (in symbols, $\Diamond \phi$, $\phi \rightarrow \psi = \Diamond \psi$).

The failure of Indicative Exclusion, we have seen, stems from the validity of substitutivity of identity for indicatives. Substitutions invoking true identity statements can turn mundane

\textsuperscript{57}There is such a possible situation, and it may well be an actual situation.

\textsuperscript{58}For [IE] implies an equivalent of [IPP]: $\Diamond \phi$, $\phi \rightarrow \psi$, $\phi \land \psi = \Diamond (\psi \land \psi)$.
It is legitimate to make substitutions in counterfactuals if they are supported by identity statements that are necessarily true (e.g., ‘$20 = $10 + $10’). But such substitutions cannot turn mundane statements to necessary falsities or compatibles to incompatibles.

Counterfactual Exclusion, for substitutivity of identity is not valid for counterfactuals. For example, we cannot appeal to (a) to substitute ‘$10’ for ‘Ali’s amount’ in the counterfactual cousin of (b):

(a) If Ali had more than $20, Ali’s amount would have been larger than $20.

Making the substitution in this statement yields the counterfactual cousin of (c):

(b*) If Ali had more than $20, $10 would have been larger than $20.

This might be false while both (a) and (b*) are true (consider situation $S^*$).

Moreover, we can give a positive argument for Counterfactual Exclusion. The thesis follows from the following:

*Counterfactual Preservation of Possibility*: If the antecedent of a true counterfactual states a possibility, its consequent must also state a possibility (in symbols, $\diamond \phi, \phi \rightarrow \psi = \diamond \psi$).

*Consequent Merging for Counterfactuals*: $\phi \rightarrow \psi$ and $\phi \rightarrow \chi$ imply $\phi \rightarrow (\psi \land \chi)$.

---

59It is legitimate to make substitutions in counterfactuals if they are supported by identity statements that are necessarily true (e.g., ‘$20 = $10 + $10’). But such substitutions cannot turn mundane statements to necessary falsities or compatibles to incompatibles.
We can intuitively see that both of these are true. Consequent merging holds for counterfactuals as well as indicatives. Although possibility preservation fails for indicatives (which we have seen), it clearly holds for counterfactuals. It might help to consider some examples to confirm this. Consider, for example, the following:

(d) Ali might have had more than $20.
(e) Ali’s amount might have been larger than $20.

(b*), together with (d), implies (e). Similarly, (c*), together with (d), implies a substitution variant of (e):

(f) $10 might have been larger than $20.

This statement, to be sure, is false, but this does not mean that the implication fails, but that (c*) must fail unless (d) does.

Counterfactual Exclusion holds, we have seen, because counterfactuals preserve possibility. The same reason does not hold for Indicative Exclusion. Indicatives do not preserve possibility, we have seen, because substitutivity of identity holds for them.

Now, note that as counterfactuals preserve possibility, indicatives preserve truth:

60Note that Lewis’s account of counterfactuals yields both theses. Similarly, Stalnaker’s account of conditionals yields their cousins that relate to indicatives as well as counterfactuals. So Stalnaker’s account yields the Exclusion Thesis, and Lewis’s the Counterfactual Exclusion Thesis.
Indicative Preservation of Truth: If the antecedent of a true indicative is true, its consequent must also be true (in symbols, \( \phi, \phi \land \psi = \psi \)).

This is simply the semantic basis of modus ponens. Similarly, indicatives and counterfactuals are governed by contrapositives of truth preservation and possibility preservation, respectively:

Backward Falsity Preservation of Indicatives: If the consequent of a true indicative is false, its antecedent must also be false (in symbols, \( \phi \land \neg \psi, \neg \psi = \neg \phi \)).

Backward Impossibility Preservation of Counterfactuals: If the consequent of a true counterfactual states an impossibility, its antecedent must also state an impossibility (in symbols, \( \phi \Box \neg \psi, \neg \Box \psi = \neg \Box \phi \)).

So I suggest that truth is to indicatives what possibility is to counterfactuals.\(^{61}\) If so, the direct counterpart of counterfactual preservation of possibility is indicative preservation of truth, not that of possibility. Indicatives, then, diverge from counterfactuals by not preserving a relative of truth,

\(^{61}\)Some might object that truth preservation holds for the usual counterfactuals as well:

Counterfactual Preservation of Truth: If the antecedent of a true counterfactual is true, its consequent must also be true (in symbols, \( \phi, \phi \Box \neg \psi = \psi \)).

But truth preservation does not hold for another sort of counterfactuals, the \( \textit{might} \)-counterfactuals: ‘Ali has more than $20’ and ‘If Ali had more than $20, it \textit{might} have had more than Baba’ do not imply ‘Ali has more than Baba.’ So I think that truth preservation holds for the \( \textit{would} \)-counterfactuals because they (unlike \( \textit{might} \)-counterfactuals) agree with indicatives when their antecedents are true. (Backward falsity preservation holds for them for the same reason.)
possibility, by being governed by substitutivity of identity. This, we have seen, gives rise to a substantial logical disparity between indicatives and counterfactuals.

7. Concluding Remarks

Smullyan’s paradox, we have seen, helps to highlight important logical differences between indicative and counterfactual conditionals. Consider two matching pairs of conditionals:

(a) If Ali has more than Baba, Ali’s amount is larger than Baba’s.

(a*) If Ali had more than Baba, Ali’s amount would have been larger than Baba’s.

(b) If Ali has more than Baba, Ali’s amount is larger than $20.

(b*) If Ali had more than Baba, Ali’s amount would have been larger than $20.

Both (a) and (a*) are true, presumably analytically, but this does not mean that there is no logical difference between them. The indicative (a) together with ‘Baba’s amount is $20’ implies (b), but the same does not hold for their counterfactual cousins, (a*) and (b*). This is because substitutivity of identity holds for indicatives, but not for counterfactuals. And this leads to a logical disparity between the following as well:

(c) If Ali has more than Baba, $10 is larger than $20.

(c*) If Ali had more than Baba, $10 would have been larger than $20.
The counterfactual (c*) is incompatible with ‘Ali might have had more than Baba’, for counterfactuals preserve possibilities: a counterfactual cannot be true if its antecedent states a possibility while its consequent states an impossibility. It is not the same with the indicative (c). This follows from (a) and ‘Ali’s amount is $10, and Baba’s is $20.’ So it can be true while its antecedent states a possibility. Similarly, the contrary indicatives ‘If Ali has more than Baba, the difference between their amounts is $5’ and ‘If Ali has more than Baba, the difference between their amounts is $10’, we have seen, can both be true while their antecedent states a possibility. This is crucial to solving Smullyan’s paradox. And it casts significant light on contemporary debates on the logic and semantics of conditionals.

Department of Philosophy
University of Toronto
Toronto, ON M5R 2M8
b.yi@utoronto.ca

---

I think it helps to solve some other puzzles, such as McGee’s objection to modus ponens for indicative conditionals, and Priest’s version of Newcomb’s problem. For the puzzles, see Vann McGee, “A counterexample to modus ponens”, Journal of Philosophy 82, 9 (September 1989): 462-471; and Priest, op.cit. For a solution to Priest’s puzzle, see my “Newcomb’s paradox and Priest’s principle of rational choice”, Analysis 63, 3 (July 2003): 237-242. I leave it for another occasion to present a solution to McGee’s puzzle.