Studies in Logic, Vol. 15, No. 6 (2022): 28–48
PII: 1674-3202(2022)-06-0028-21

Counterfactual Triviality and Structural Causal Models*

Xiaoan Wu

Abstract. J. Williams (2012) proposes a version of counterfactual triviality (CT). I carefully examine the four premises on which Williams' CT relies. Within the framework of the Structural Causal Model (SCM), I show that two of them (CRT and CVPP) apply to two different types of counterfactuals respectively, so that the $P^A(C)$ of them is not equivalent, thus proving that Williams' version of CT is not valid.

1 Introduction

In the discussion of conditionals, the following two categories are generally involved: *indicative* conditionals and *counterfactual* conditionals. Although the distinction between these two types of conditionals is not so clear in Chinese (at least syntactically), in English there is a clear grammatical form to distinguish these two types of conditionals. There is a lot of discussion about the connections and differences between them. First of all, the semantics of these two types of conditionals are indeed distinct. For example, if it is known *Journey to the West* was actually written by Wu Cheng'en. If we "indicatively" suppose Wu Cheng'en did not write the book, then it can be assumed that someone else wrote it; But if we "subjunctively" suppose Wu Cheng'en hadn't written the book, it is likely nobody would have written the book.

Second, these two types of supposition play different roles in different life situations. In hypothesis testing and confirmation, indicative suppositions play an important role. If evidence E supports hypothesis H, then E is more likely to occur under the indicative supposition H than the indicative supposition $\neg H$. Counterfactual supposition is important in many areas, including decision-making, blame, explanation,

Received 2022-09-29 *Revision Received* 2022-11-14

Xiaoan Wu School of Marxism, Northwestern Polytechnical University
Major Public Information Research Center of Shaanxi Province
wuxiaoan1984@126.com

*This research was supported by the National Social Science Fund Subsequent Funding Project of China (21FZXB040), the General Project of MOE (Ministry of Education) Foundation on Humanities and Social Sciences (21XJC720002) and Key (Key grant) Project of The National Social Science Fund of China (17ZDA024). The author is grateful to professor Jiji Zhang of Hong Kong Baptist University for bringing this issue to my attention, and even more so for the great help he gave me in discussing this issue.

and diagnosis. When it is counterfactually assumed that a person did not set the fire, we can infer that the fire would not have occurred and we would judge that the person is inescapably responsible for the fire and should be punished by law.

Finally, just as it is debated whether or not indicative conditionals have truth values (for example, Adams in [1] argues indicative conditionals do not have truth values.), there is a debate over whether counterfactual conditionals are truth-valued propositions. For instance, Edgington ([7]) argues that counterfactuals are simply expressions of our belief-modification strategies. Some people believe that counterfactuals are only assertable, acceptable, or probabilistic, but have no truth value. But intuitively we believe or accept a counterfactual conditional with its truth in mind. Thus Hájek ([10]) believes if a counterfactual has no truth value, it is difficult to understand how a counterfactual is probabilistic. Convinced by Hájek's arguments, let's accept the assumption that the counterfactuals have truth value and continue our discussion.

So the next question is: to what degree should we believe in a conditional? There is an admittedly fairly reasonable restriction on the degree of conditionals, commonly known as the *Ramsey test*:

> If two people are arguing 'If p, then 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; so that in a sense 'If p, q ' and 'If $p, \neg q$ ' are contradictories. We can say that they are fixing their degree of belief in q given p. If p turns out false, these degrees of belief are rendered void. If either party believes not p for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses. ([28], p. 143)

The above statements are still unclear. First, what does it mean to "add p hypothetically to their stock of knowledge"? As noted above, there are at least two different kinds of supposition, one that is an indicative supposition and the other is a counterfactual supposition. Different suppositions correspond to two different types of conditionals in natural language, counterfactual conditionals and indicative conditionals, and different types of conditionals also correspond to different degrees of belief, that is, through the indicative supposition, we determine our credence for indicative conditionals, and through the subjunctive supposition, we determine credence for a counterfactual.

Second, how should we characterize and represent the above "adding"? The Ramsay test is generally understood as the conditions under which we can reasonably believe a conditional. That is, if he accepts the consequent under the supposition that the antecedent holds, then he should accept the conditional. In the framework of credence or subjective probability, it can be restated as: One's degree of rational belief in a conditional $p \to q$ should equal one's degree of credence in q, under the supposition that p. Different understandings of the above suppositions and credence have in turn produced different ways to characterize and represent the above "adding", resulting in different versions of the Ramsay test.

Our credence in an indicative conditional $A \to B$ should equal one's credence in the consequent B on the indicative supposition of its antecedent A. That is, the Indicative Ramsey Test (IRT), formally expressed as:

$$P_A(B) = P(A \to B) \tag{1}$$

The above formula links the conditional to the credence. Intuitively, the above formula is acceptable. If it is assumed that Wu Cheng'en did not write *Journey to the West* (that is, Wu Cheng'en did not actually write *Journey to the West*), then there is a high probability that *Journey to the West* was written by someone else, and likewise, you have a high probability of accepting the conditional "If Wu Cheng'en did not write *Journey to the West*, then someone else wrote the book".

Our credence in a counterfactual conditional $A \mathrel{\Box\!\!\to} B$ should equal one's credence in the consequent B on the subjunctive supposition of its antecedent A. That is, the Counterfactual Ramsey Test (CRT), formally expressed as:

$$P^A(B) = P(A \mathrel{\Box\!\!\to} B) \tag{2}$$

In fact, the possible world semantics given by Stalnaker ([36]) and Lewis ([16]) to determine the truth value of a counterfactual is consistent with the idea of CRT. According to Lewis, the truth value of the counterfactual $A \mathrel{\Box\!\!\to} B$ in the actual world is determined by the truth value of B in the A-world closest to the actual world.

But if introducing credence and accepting that a counterfactual's truth value affects our credence in it, then CRT does not seem to hold. For example, considering the counterfactual "If I had flipped this fair coin, it would have landed heads ($A \mathrel{\Box\!\!\to} B$)", assuming the antecedent is true, and the confidence in the consequent is 50%. But according to their semantics, because there does not exist an $A \wedge B$-world closer to the actual world than any $A \wedge \neg B$-world, so the counterfactual $A \mathrel{\Box\!\!\to} B$ is false, then our credence in it is relatively low, less than 50%, so CRT is not valid.

Given that in our subsequent discussion of CT, the possible world semantics of counterfactuals are not a premise and basis we have to adhere to, so let us accept CRT for the time being, which is intuitively reasonable. If you counterfactually suppose that Wu Cheng'en hadn't written *Journey to the West*, and you think there is a high probability that nobody would have written the book *Journey to the West*, then it seems equally reasonable to agree that the following counterfactual has a great probability:

If Wu Cheng'en hadn't written *Journey to the West*, then nobody would have written the book.

Third, on standard Bayesian construals, credence under indicative supposition is identified with conditional probability. So there is one interpretation of the Ramsay test: Adams' Thesis (AT), also known as Stalnaker's Thesis (ST), formally expressed as:

$$P(A \to B) = P(B \mid A) \tag{3}$$

Note that although AT and ST have the same form, the interpretations of $P(A \to B)$ are not the same. And let us denote the different interpretations of $P(A \to B)$ by Adams and Stalnaker by $P'(A \to B)$ and $P^*(A \to B)$ respectively. Adams ([2]) thinks that $P'(A \to B)$ can be understood as expressing "the assertability of $A \to B$"; And Stalnaker ([37]) argues that $P^*(A \to B)$ should be understood as "the probability that $A \to B$ is true", which is equivalent to the probability $P(B \mid A)$. These different interpretations have their own validity and do not affect the next discussion, so let's take a neutral stance on them.

Lewis ([17]) pointed out that no matter how $P(A \to B)$ is understood, as long as P obeys the laws of probability, then the Adams thesis plus the possible world semantics of conditionals yields the following triviality results: [1]

$$P(A \to B) = P(B) \tag{4}$$

This triviality result implies that in non-trivial language, AT cannot be accommodated within the framework of classical possible-world semantics, and Bradley ([4]) also shows that even a very weak consequence of AT is not compatible with the framework of classical possible-world semantics. For a discussion of triviality problems in Chinese see Su ([38]) and Liu ([23]).

[1] The specific proof process is as follows:

$$\begin{aligned}
P(B \mid A) &= P(A \to B) \\
&= P(A \to B \mid B)P(B) + P(A \to B \mid \neg B)P(\neg B) \\
&= P'(A \to B)P(B) + P''(A \to B)P(\neg B) \\
&= P'(B \mid A)P(B) + P''(B \mid A)P(\neg B) \\
&= P(B \mid A \wedge B)P(B) + P(B \mid A \wedge \neg B)P(\neg B) \\
&= P(B)
\end{aligned}$$

This proof's validity is based on two assumptions. First, $P(A \to B \mid B) = P'(A \to B)$ implies the assumption that $\to$ denotes the same propositional connective in any context. If P' obtained by conditionalizing P on the proposition B (P'' obtained by conditionalizing P on the proposition $\neg B$), then this conditionalisation does not affect the truth condition of $P(A \to B)$. Second, $P'(A \to B) = P'(B \mid A)$ implies the assumption that the Adams thesis holds not only for P, but also for P'.

As noted above, in the case of indicative conditionals, there are standardized ways to formalize the credence under the indicative suppositions — i.e., by conditional probabilities. In the case of counterfactual conditionals, there is no universally accepted standard way of implementing the subjunctive suppositions, but rather a variety of positions, none of which is universally accepted. Often referred is the position of Skyrms ([35], p. 261). In a critique of Adams ([3]), he replaced "prior epistemic probabilities" with "a priori propensities", so the **Skyrms' Thesis** (ST) equates credence in the counterfactual with the expectation of the corresponding conditional chances:

$$P(A \mathrel{\Box\!\!\rightarrow} B) = \sum_i Ch_i(B \mid A) \cdot P\left(CH_i\right) \tag{5}$$

where Ch_i is the objective chance function and CH_i says that Ch_i is correct about the objective chance, and further assumes that all CH_i (denoted as $\{CHi\}$) are a partition of chance hypotheses.

Although very often we do not know the objective chance of a specific event or proposition, ST implies that when the objective chance of the consequent is above a certain threshold given the antecedent, then we should accept the counterfactual. Moreover, the reason for the weighted expression in equation (5) above is that we are not sure about the exact value of the conditional chance, so we have made a partition of the possible chance propositions. Finally, whether ST is true is still controversial and not generally accepted, and there are alternative ways of cashing out subjunctive supposition ([31]). The above discussion can be summarized in the following figure:

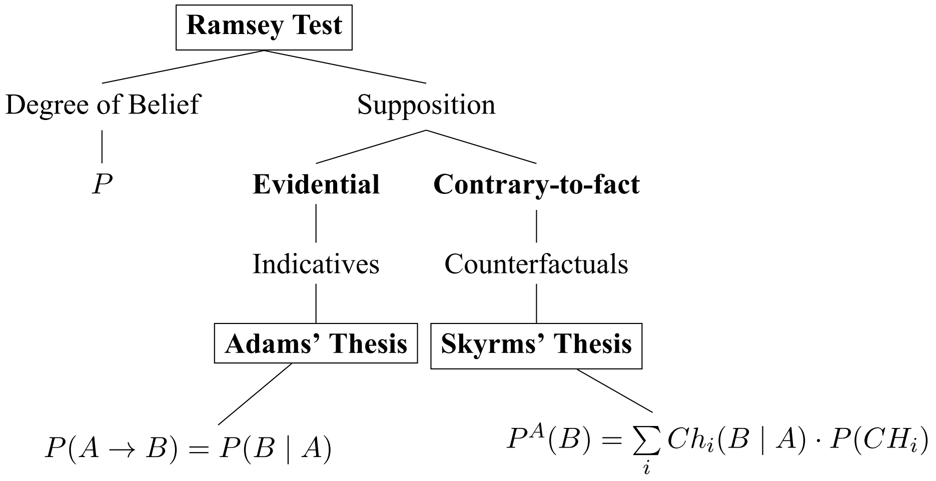


Figure 1: The Origins of Adams' Thesis and Skyrms' Thesis

Finally, Although the triviality of indicative conditionals has been much discussed, few have heard of CT , and Williams ([39]) demonstrates that under some seemingly reasonable assumptions, we can derive CT results as follows:

$$P(A \mathrel{\Box\!\!\rightarrow} B) = P(B) \tag{6}$$

First, this is obviously a very strange and absurd result. For example, consider a counterfactual conditional in Chinese: "*If only the Winged General of Han were around to fight the township of Basilisk, the barbarians and their horses would never have dared to cross the Mountains of Yin*". If the above triviality holds, it means that we have the same degree of belief in the proposition "*If only the Winged General of Han were around to fight the township of Basilisk, the barbarians and their horses would never have dared to cross the Mountains of Yin*" as we have in the proposition that "*The barbarians and their horses would never have dared to cross the Mountains of Yin*".

Second, for researchers working with SCM and potential outcomes models, CT is a very strange result, and it seems that no one has had time to think about what such a result means for SCM, this seems to be a quirk that only arises in logical contexts. If this is right, it is clearly a problem that needs to be addressed. But for now, most people are skeptical of this result ([5, 33]), and this paper also tries to show that this result is not true from the perspective of SCM.

2 Williams' Argument

Next, we will discuss specifically how Williams derived the triviality results. It is important to note that CT actually takes many different forms, depending on different presuppositions. Triviality results derived by Williams ([39]) are closely related to four *prima facie* reasonable premises, in particular, **the Principal Principle** (PP), which links chance and credence, and ST (i.e., A Conditional Version of the Principal Principle, CVPP), while other versions of the CT results are not based on these premises. For example, Santorio ([30]) uses six plausible hypotheses "Nonzero, Upper Bound, CRT, Restricted Suppositional Additivity, CNC and Closure" to obtain the triviality result: $P(A \mathrel{\Box\!\!\rightarrow} B) = P(A \mathrel{\Diamond\!\!\rightarrow} B)$. This paper will focus on the triviality results given by Williams ([39]) and give an interpretation of the triviality results of Santorio ([30]) based on my solution to Williams' triviality results.

Broadly speaking, Williams' argument can be structured as below.

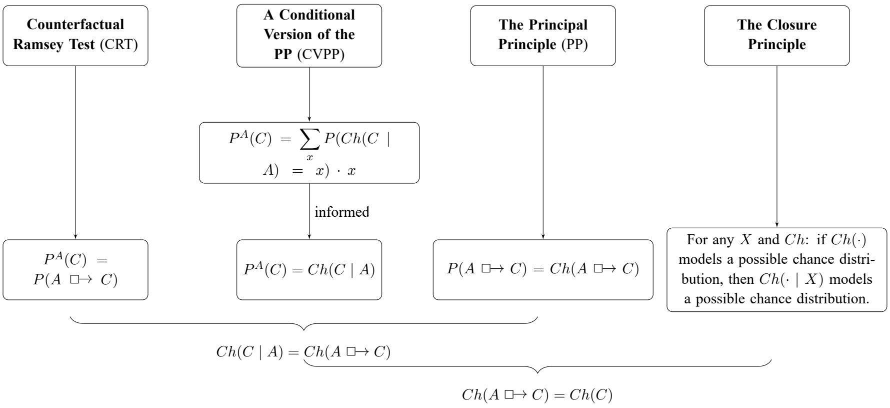

Counterfactual Ramsey Test (CRT)
A Conditional Version of the PP (CVPP)
The Principal Principle (PP)
The Closure Principle
$P^A(C) = \sum_x P(Ch(C \mid A) = x) \cdot x$
informed
$P^A(C) = P(A \boxright C)$
$P^A(C) = Ch(C \mid A)$
$P(A \boxright C) = Ch(A \boxright C)$
For any X and Ch: if $Ch(\cdot)$ models a possible chance distribution, then $Ch(\cdot \mid X)$ models a possible chance distribution.
$Ch(C \mid A) = Ch(A \boxright C)$
$Ch(A \boxright C) = Ch(C)$

Figure 2: Williams' Argument.

In the discussion that follows, we will analyze each of these four premises and show how they ultimately derive the triviality results.

First, as Williams ([39], p. 649) states, the CRT is just a normative constraint[2], which means that for a fully rational agent, his categorical credence in the counterfactual $A \boxright B$ and his degree of belief in B on the counterfactual supposition that A should coincide. There may be cases in which it is not satisfied, but if this normative constraint is correct, then the above violation is "a form of irrationality".

Second, another important premise of Willams' argument is PP. Lewis ([21], p. 266) proposed PP (Slightly modified for consistency with the symbolic expressions in this paper).

> **PP**. Let P be any reasonable initial credence function. Let t be any time. Let x be any real number in the unit interval. Let X be the proposition that the chance, at time t, of A 's holding equals x. Let E be any proposition compatible with X that is admissible. at time t. Then
>
> $$P(A \mid X, E) = P(A \mid Ch_t(A) = x, E) = x$$

Looking at this definition, you may find the principle very complex and involves many concepts that need further clarification, so let's start with a simple example of

[2] In Williams ([39]), he discusses not only the CRT (also known as his *Counterfactual Ramsey Identity*), but also alternative assumptions, such as *Counterfactual Ramsey Bound* and *Counterfactual Ramsey Zero* respectively, those premises also lead to absurd results. It seems that these absurd results are all based on the above assumptions and the equivalence of the $P^A(C)$ in ST and CRT, so this paper's refutation of the argument premised on the CRT also constitutes a refutation of the other two alternative assumptions.

the application of the principle so as to get an intuitive grasp of the principle: Suppose you are going to throw a die and you want to assign your degree of belief to various assumptions about the number of dice, what should be your degree of belief about the number of dice of 3? According to Lewis ([19]), your belief in it is determined by PP. Let P denote your subjective credence, A is the proposition that the number of dice tossed is 3, $Ch(A) = x$ is the proposition that the objective chance of A is x, and E is a proposition that must be *admissible*.

So PP essentially says that, in the absence of *inadmissible* information, your beliefs about the chance of the dice being thrown at 3 should guide your beliefs about the dice being thrown at 3:

$$P(A \mid Ch(A) = x, E) = x.$$

Lewis does not give a precise definition of what *admissible* information is and what *inadmissible* information is. Lewis ([21], p. 272) says roughly: “Admissible propositions are the sort of information whose impact on credence about outcomes comes entirely by way of credence about the chances of these outcomes.” And he also gives two examples of what is generally admissible information: *historical information* and *hypothetical information about chance itself.* The ripple effect of Lewis’s question about admissible information (and the controversy it generated) continues to this day. To take our current example, inadmissible information means that, before you throw the dice, if an omniscient prophet tells you that your next throw will be a 3, then as a rational person, your confidence that the number of dice will be 3 will be greatly increased, even though your belief in the objective chance of the number of dice being 3 has not changed. Therefore, the discussion of the relationship between chance and credence requires the absence of inadmissible information, otherwise PP would not hold.

So without inadmissible information, according to PP, we have:

$$P(A \mathbin{\Box\!\!\rightarrow} C) = Ch(A \mathbin{\Box\!\!\rightarrow} C).$$

Third, to understand the conditional version of the Principal Principle (CVPP), we have to start with *causal decision theory* (CDT). As Joyce ([15], p. 161), “Causal decision theory seeks to provide a rigorous formal analysis of the idea that a rational decision maker should evaluate her potential actions solely based on their ability to cause desirable outcomes.” We can formalize this idea as a function of causal expected utility in the following form:

$$U(A) = \sum_{C} P^{A}(C)\, u(A \,\&\, C) \qquad \textbf{(CDT)}$$

The probability function $P^{A}(\cdot)$ measures the agent’s estimate of the ‘causal tendencies’ of A; $U(A)$ measures the extent to which performing of A can lead to desirable or undesirable outcomes; and the function $u(\cdot)$ represents the desirability of the agent for various states of the world. According to CDT, the agent should choose an action that maximizes the causal expected utility.

The most crucial aspect of the above equation is interpreting $P^A(C)$, which has been interpreted differently by different causal decision theorists, but as Joyce ([15], p. 161) points out, all of them rest on a common foundation: "$P^A(C)$'s values should reflect a decision maker's judgments about her ability to *causally influence* events in the world by doing A. I will call it her *causal probability for A.* "

Lewis ([20], p. 11) proposed a $\mathbb{K}$-partition account of causal probabilities, where an element K in a particular partition $\mathbb{K}$ is "a maximal specific proposition about how what [the agent] cares about depends causally on his present actions". Like Lewis, we call the elements in $\mathbb{K}$ the *dependency hypotheses*. Thus, we have a measure of the 'causal tendencies' or causal probability for A ([15], p. 164):

Definition 1 ($\mathbb{K}$-expectation Definition of Causal Probability). If $P(A) > 0$, then

$$P^A(C) = \sum_{K \in \mathbb{K}} P(K)\, P(C \mid A \,\&\, K) \tag{7}$$

for some appropriate choice of a partition of dependency hypotheses $\mathbb{K}$.

The next problem is to find the appropriate $\mathbb{K}$. One interpretation of $\mathbb{K}$ claims that dependency hypotheses provide direct specifications for objective chances. every K contains a complete theory of objective conditional chance such that for each event C and act A, it implies a proposition of the form $Ch(C \mid A) = x$. Thus, in the context of decision theory, we can assume that conditional chance and subjective probability are related as follows:

Definition 2 (CVPP). If P is any probability on Ω, if $P(A, K) > 0$, and if K entails that the chance of C conditional on A is x, then

$$P(C \mid A, K) = P(C \mid A, Ch(C \mid A) = x) = x. \tag{8}$$

According to CVPP, the equation can be further decomposed as follows:

$$\begin{aligned} P^A(C) &= \sum_K P(K)\, P(C \mid A, K) \\ &= \sum_x P(Ch(C \mid A) = x) \cdot x \end{aligned}$$

The result obtained above is *ST* (cf. Eq.5) ! Consider an agent who is fully informed about Ch, the above equation simplifies to:

$$P^A(C) = Ch(C \mid A)$$

Fourth, *The Closure Principle* states that if $Ch(\cdot)$ models a possible probability distribution for any proposition X and chance function $Ch(\cdot)$, then $Ch(\cdot \mid X)$ also models a possible probability distribution. As with "Conditional probabilities are probabilities" , there does not seem to be much hesitation in accepting this principle.

Based on the above four premises, Williams ([39], p. 661) derives:

$$\begin{aligned} Ch(A \mathrel{\Box\!\!\rightarrow} C) &= Ch(A \mathrel{\Box\!\!\rightarrow} C \mid C)\, Ch(C) + Ch(A \mathrel{\Box\!\!\rightarrow} C \mid \neg C)\, Ch(\neg C) \\ &= Ch'(A \mathrel{\Box\!\!\rightarrow} C)\, Ch(C) + Ch''(A \mathrel{\Box\!\!\rightarrow} C)\, Ch(\neg C) \\ &= Ch'(C \mid A)\, Ch(C) + Ch''(C \mid A)\, Ch(\neg C) \\ &= 1 \cdot Ch(C) + 0 \cdot Ch(\neg C) \\ &= Ch(C) \end{aligned}$$

First, like indicative triviality, Ch' obtained by conditionalizing Ch on the proposition C, $Ch(A \mathrel{\Box\!\!\rightarrow} C \mid C) = Ch'(A \mathrel{\Box\!\!\rightarrow} C)$ implies that "$\Box\!\!\rightarrow$" expresses the same propositional connective in every context, so the conditionalisation do not affect the truth condition of $Ch(A \mathrel{\Box\!\!\rightarrow} C)$, Williams does not specifically discuss the legitimacy of this assumption, obviously this assumption can be challenged ([29]), but it is not the subject of this paper, so let's put it aside. Second, according to CVPP, PP and CRT, we have $Ch'(A \mathrel{\Box\!\!\rightarrow} C) = Ch'(C \mid A)$; Third, according to PP, we can finally obtain equation (6), which yields the CT results.

3 Contra Counterfactual Triviality

Although this result is unacceptable, the four principles on which it is based seem reasonable. "Conditional probability is also a probability" is a proven probability theorem, so the Closure Principle seems feasible. PP (and CVPP) is also an intuitively compelling principle, although Lewis ([22], p. 473) turns to a more complex "New Principle" because of the "one big and bad bug", the bug is problematic because of his Humean Supervenience conception of the nature of chance, the plausibility of which is not in question if the ontological position of Humean Supervenience is put aside. Schwarz ([32]) also gives a formal proof of PP. As mentioned before, if we hold possible worlds semantics of counterfactuals, then CRT does not hold. But given the many problems with the possible worlds semantics itself, and the fact that it can be discussed without presupposing any semantics, CRT is not problematic as a normative principle.

So each of these principles has its own focus and scope of application. First, CRT considers the conditions under which a rational agent can reasonably believe a counterfactual, or assign probabilities to counterfactuals. Note that this normative constraint does not conflict with the fact that counterfactuals are context-dependent. It is well-known that the counterfactuals $A \,\Box\!\!\rightarrow B$ are context-dependent ([34], pp. 257–259), so it is reasonable to assume that $P(A \,\Box\!\!\rightarrow B))$ is also context-dependent. A counterfactual sentence can be interpreted in multiple ways depending on the conversational context, intention and practical purpose of the speaker at the time. So for the counterfactual contextualist, the truth value of a counterfactual depends on what the speaker is trying to say when he says the counterfactual and what the hearers think the speaker is saying when he hears the counterfactual. In one context, "If Caesar had been in charge [in Korea], he would have used the atom bomb" is true, while in another context "If Caesar had been in charge [in Korea], he would have used catapults" can also be used with a high degree of belief. But in any context, credence in the counterfactual is governed by normative constraints like CRT.

Second, the objective chance is independent of the above contextual factors, and our everyday understanding of chance is consistent with the physics understanding of chance. Chance is a characterization of the objective features of the world, not a characterization of the uncertainty of an agent. The chance of a tritium atom decaying in 2023 is clearly not relevant to the context of the conversation. When A and C are context-independent, the conditional odds $Ch(C \mid A)$ are also context-independent, according to the usual analysis:

$$Ch(C \mid A) = \frac{Ch(A,C)}{Ch(A)} \quad (\text{ if } Ch(A) > 0)$$

Both $Ch(A,C)$ and $Ch(A)$ are objective chances with fixed content, so they are context independent, and their ratios are also context-independent. Thus, the conditional chance $Ch(C \mid A)$ is context independent. Therefore, a direct refutation of Williams' argument is that the $P^A(C)$ used in the CRT and CVPP is not equivalent. The former is compatible with the context-dependent fact of the counterfactual, while the latter abandons the context-dependence of the counterfactual altogether, but many counterfactuals are under-described, e.g., "If Caesar had been in charge [in Korea]" and "If a chicken had lips"([13], p. 1165). And without the addition of precise information, or in the absence of precise context, there is no objective chance of the above counterfactual.

In the following discussion, I will further illustrate that $P^A(C)$ in CRT and $P^A(C)$ in CVPP are not equivalent within the framework of SCM. SCM ([25]) is a methodological model of social science developed by computer scientist Judea Pearl and his disciples. Unlike the thinking of philosophers and logicians, he is an application-oriented scientist. The focus of his thinking is not on examining universal principles and their legitimacy for counterfactuals or causation, but rather on how to construct models to answer specific counterfactual and causal inference questions. And their research is useful for philosophical and logical thinking about counterfactuals in general. This general thinking can easily lead us to ignore possible differences between specific counterfactuals and to claim premises or principles that are intuitively reasonable but not actually true, leading to absurd results. Referring to the SCM for the distinction and solution of specific counterfactuals and causal inference problems, or the algorithmic implementation of specific counterfactuals, will allow us to better understand the differences and distinctions between the counterfactuals themselves, thus clarifying the boundaries of the applicability of some general principles (e.g., CRT and CVPP), and thus dissipating the absurd results.

The following argument will be divided into four steps: First, I will prove CRT $\Leftrightarrow P^{X=x}(Y=y) \Leftrightarrow P(X=x \mathrel{\Box\!\!\rightarrow} Y=y) \Leftrightarrow P(Y_x=y \mid x', y')$; Second, prove CVPP$\Leftrightarrow$ $P^{X=x}(Y=y) \Leftrightarrow Ch(Y=y \mid X=x) \Leftrightarrow P(Y=y \mid do(X=x))$; Third, prove $P(Y_x=y \mid x', y') \neq P(Y=y \mid do(X=x))$; Fourth, I proof that the CT is not established.

The key to the proof is to recognize that CRT and CVPP as normative constraints actually correspond to two different types of counterfactuals: *retrospective* counterfactuals and *prospective* counterfactuals, although this distinction has not yet been made explicit (These two types of counterfactuals respectively correspond to the second-level "intervention" and the third-level "counterfactual" in the Three-Level Causal Hierarchy given by Pearl). For example, Hitchcock ([12], p. 130) points out that the causation involved in CDT is not actual causation in the Lewisian sense: "What is distinctive about actual causation is rather that is retrospective: it involves a kind of reasoning backward from effects to their causes. By contrast, CDT is prospective: it involves reasoning forward from causes to their effects."

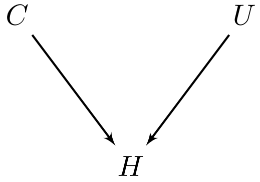
Similarly, Pearl, Glymour, and Jewell ([27], p. 90) have argued that when driving home from work and passing a fork in the road, the inference made by choosing to take one of the roads is different from the inference made by taking that road and then counterfactually imagining what would happen if you took the other highway: "My retrospective estimate is that a freeway drive would have taken less than 1 hour, and this estimate is clearly different than my prospective estimate was, when I made the decision prior to seeing the consequences—otherwise, I would have taken the freeway to begin with." Given the close connection between causation and counterfactuals, the counterfactuals involved in CRT are those corresponding to actual causation, i.e., counterfactuals in the subjunctive mood, whereas CVPP is derived from CDT, and its counterfactuals are interventionist counterfactuals, which can be understood in the indicative mood.([14])

3.1 The algorithmization of counterfactuals

Within the framework of SCM, the calculation of the retrospective counterfactual follows a fixed pattern, which can be illustrated by the following paradigmatic counterfactual:

(a) If I had flipped this fair coin, it would have landed heads.

First, for the representation of the world, SCM has a Laplace's quasi-deterministic conception that all randomness is due only to unknown causal factors, these known and unknown causal factors together constitute a deterministic conception of causation. For example, the randomness in the coin flip example is represented by the variable U:

Causal Diagram	**Structural Equations**
	$H = C \times U$

$$\begin{cases} H = 0 & \text{no result} \\ H = 1 & \text{head} \\ H = 2 & \text{tail} \end{cases} \quad \begin{cases} C = 0 & \text{no flip} \\ C = 1 & \text{flip} \end{cases} \quad \begin{cases} U = 1 & \text{unknown causal factors} \\ U = 2 & \text{unknown causal factors} \end{cases}$$

$$P(U = 1) = 1/2 = P(U = 2)$$

Second, within the framework of SCM, the calculation of the counterfactual probability goes through three steps: Abduction, Action, and Prediction. The basic idea behind this is actually very simple. In order to correctly state what will happen under the counterfactual supposition, it is necessary to have an exact grasp of the real situation. therefore, a fully specified model is needed. Abduction is based on the known results to determine the specific background, then the counterfactual antecedent is realized by intervention, and finally the probability of the consequent is calculated under the condition that the antecedent occurs, and thus the probability of the counterfactual is finally determined.

For example, the probability of (a) can be calculated according to Pearl ([25], p. 206). According to Abduction, we obtain:

$$P(u \mid C = 0, H = 0) = \begin{cases} 1/2 & u = 1 \\ 1/2 & u = 0 \end{cases}$$

According to action and prediction, we get:

$$\begin{aligned} P(C = 1 \mathrel{\Box\!\!\rightarrow} H = 1) &= P(H_{C=1} = 1 \mid C = 0, H = 0) \\ &= P(U = 1 \mid C = 0, H = 0) \\ &= 1/2 \end{aligned}$$

The probability of counterfactual (a) is 1/2, which fits with the intuition. So within the framework of SCM, we can have the following conclusions

$$\begin{aligned} P^{X=x}(Y=y) &= P(X=x \mathrel{\Box\!\!\rightarrow} Y=y) \\ &= P(Y_x = y \mid x', y') \end{aligned}$$

3.2 Causation decision theory

In the previous discussion on CDT, we discussed a $\mathbb{K}$-partition account of causal probabilities for A (Eq. (7)), and it is known that the most difficult part of the theory here is how to understand $\mathbb{K}$. Eq. (8) gives a solution: the "chance" reading of dependency hypotheses $\mathbb{K}$, and there is another way of understanding belief about causal tendencies, namely the "counterfactual dependence" reading of dependency hypotheses:

> We begin with a rough theory of rational decision-making. In the first place, rational decision-making involves conditional propositions: when a person weighs a major decision, it is rational for him to ask, for each act he considers, what would happen if he performed that act. It is rational, then, for him to consider propositions of the form 'If I were to do a, then c would happen'. Such a proposition we shall call a *counterfactual*. ([9], p. 153)

So a function of causal expected utility can be expressed as follows:

$$U(A) = \sum_{C} P(A \mathrel{\Box\!\!\rightarrow} C)\, u(A \,\&\, C) \tag{9}$$

The term '$\Box\!\!\rightarrow$' refers to non-backtracking counterfactuals in the sense of Lewis ([18]), but since a formal method for combining chance and counterfactuals is missing, it is not clear how to compute $P(A \mathrel{\Box\!\!\rightarrow} C)$, though the idea is there.

Meek and Glymour ([24]) pointed out that we can elaborate $P(A \mathrel{\Box\!\!\rightarrow} C)$ using the formalism of doing interventions in Bayesian networks. Hitchcock ([13]) further demonstrates that this proposal not only helps to clarify a number of issues surrounding CDT, but also constitutes a response to many of the "exotic" counterexamples to this theory.

Although we use "counterfactuals" here, it is important to note that counterfactuals here are different from the counterfactuals mentioned in the previous section that correspond to actual causation, and are a special class of counterfactuals, as Edgington says:

> Note: I shall stick to the label "counterfactual", as most participants in the debate do, because the issue is not really one of grammar but one of function. But, not to be misled, you have to realize that these are "counterfactuals" which do not presuppose the falsity of the antecedent. It is just a convenient label for a type of conditional, a conditional which in English has a "would" in the consequent, which includes those that do presuppose the falsity of the antecedent. ([8], p. 78)

DeRose ([6]) further states that although it is widely assumed that 'straightforward' future-directed conditionals that are used in CDT are counterfactual or subjunctive conditionals, he argues that the conditionals of deliberation are **indicative**. We do not intend to discuss in depth whether the counterfactuals used in decision theory are paradigmatic counterfactuals or indicative conditionals, but at least the counterfactuals involved in decision theory are a special class of counterfactuals, as envisioned by Meek and Glymour ([24]), using Pearl's do-operator, in the SCM, as given by Pearl ([26], p. 981), one can further express equation (9) as:

$$U(A) = \sum_{C} P(C \mid do(A))\, u(A \,\&\, C) \tag{10}$$

In summary, we have given two understandings of $P^A(C)$, one is the "chance" reading and the other is the "counterfactual dependence" reading (or do-operator), as Lewis says: "We causal decision theorists share one common idea, and differ mainly on matters of emphasis and formulation. "([20], p. 5) Harper and Skyrms also says:"It can be argued that the various forms of CDT are equivalent — that an adequate version of any one of [them] will be interdefinable with adequate versions of the others."([11], p. x) If the above understanding is correct, then:[3]

$$\begin{aligned} P^{X=x}(Y=y) &= Ch(Y=y \mid X=x) \\ &= P(Y=y \mid do(X=x)) \end{aligned}$$

The anonymous reviewer questions the legitimacy of the above equation, "since CDT is a decision theory, the probabilities involved in $U(A)$ must be subjective probabilities, but $Ch(Y = y \mid X = x)$ is an objective probability, and they cannot be equal", while in my understanding, the $P(Y = y|do(X = x))$ used in $U(A)$ is a representation of the interventionist counterfactual probabilities. According to the SCM, when there is an accurate causal diagram characterizing the specific situation and sufficient reliable data, we get the objective causal effect of the antecedent on the

[3] "As philosophers of science have long been telling us, the notions of causation, chance, counterfactual dependence, similarity among worlds, and natural law form a constellation of interrelated concepts, any one of which can be used as a starting point for an analysis of the rest." ([15], pp. 171–172)

consequent, while $Ch(Y = y \mid X = x)$ characterizes the objective probability of the consequent $(Y = y)$ occurring if the antecedent $(X = x)$ occurs. So it seems to me that, first of all, what they characterize is actually the same. Second, as Joyce points out, "As philosophers of science have long been telling us, the notions of causation, chance, counterfactual dependence, similarity among worlds, and natural law form a constellation of interrelated concepts, any one of which can be used as a starting point for an analysis of the rest. "([15], pp. 171–172) So although $u(A\&C)$ in $U(A)$ represents the desirability of the subjective agent, $P(Y = y \mid do(X = x))$ characterizes the objective causal effect of the occurrence of the antecedent on the consequent.

3.3 Inequivalence in the SCM

Next we have to prove that:

$$P(Y_x = y \mid x', y') \neq P(Y = y \mid do(X = x))$$

Note that $\neq$ here is not an 'inequality' in the mathematical sense, but rather that the two are not reciprocal reductions. In the framework of SCM, this inequivalence is obvious.[4]

As mentioned earlier, $P(Y_x = y \mid x', y')$ is a formal characterization of the credence in the retrospective counterfactual, while $P(Y = y \mid do(X = x))$ (or $P(Y_x = y)$) is primarily a formal characterization of the credence in the prospective counterfactual. First, the difference can be illustrated from the perspective of the possible worlds, $Y_x = y$ and $(X = x', Y = y')$ in $P(Y_x = y \mid x', y')$ are events that occur in different possible worlds, $(X = x', Y = y')$ in the real world, and $Y_x = y$ in the counterfactual world in which $X = x$ holds. In order to determine the value of Y_x in this counterfactual world, we need the information from the real world: $(X = x', Y = y')$; while $(X = x, Y = y)$ in $P(Y = y \mid do(X = x))$ occurs in the real world and does not involve the counterfactual world.

Second, the causal issues explored by the two representations are not the same. Using $P(Y_x = y \mid x', y')$ is more about the **Causes of Effects** (CoE). For example, to determine whether receiving irradiation was the cause of the patient's tumor recurrence (actual causation) in a realistic scenario where the patient did not receive irradiation and his tumor recurred, one must examine the credibility or truth of the counterfactual: "If I had gone through irradiation, my tumor would not have recurred."

[4] When I discussed my thinking with professor Jiji Zhang, Jiji pointed out that if Williams asserts that the counterfactuals he discusses are interventionist counterfactuals, then my rebuttal is no longer viable. I agree with what Jiji says, but as far as I understand it, I do not think Williams would take the position that the scope of CRT as he understands would not include the paradigmatic counterfactuals discussed in §3.1, because, first of all, such counterfactuals are too common to be ignored, and secondly The discussion of counterfactuals by philosophers has also focused mostly on this.

Because CDT is about predicting the outcome of each action option causally, the use of $P(Y = y \mid do(X = x))$ is more about **Effects of Causes** (EoC). For example, in a realistic situation, the patient has not yet received any treatment and has two choices in front of him, either irradiation or no irradiation. The patient has to make predictions about the 'causal tendencies' of these two actions, and then choose the action that maximizes the causal expected utility based on the desirability of the various outcomes, by examining the truth or belief of the following counterfactual: "If I were to receive irradiation, then my tumor would not recur." In the EoC quest, the potential actions under study are chosen ahead of time, whereas, in the CoE quest, the research goal is to find and access the importance of causes. From an experimentalist perspective, $Y_x(u)$describes the behavior of a specific individual $U = u$ under the intervention $do(X = x)$ (or, of course, the behavior of a sub-population). So we can use this formal picture as a basis for discussing some ethical concepts: credit, blame, and regret. But $P(Y = y \mid do(X = x))$ characterizes the behavior of a population under a given intervention.

Third, the 'equipment' and methods needed to calculate the two are different. To compute the exact probability of the counterfactual, we need data and a fully specified model, Pearl ([25], p. 206) gives three steps for the computation: Abduction, Action, and Prediction. But to compute $P(Y = y \mid do(X = x))$, we just need the data and a causal diagram that correctly articulates 'the story behind data — the causal mechanism that led to, or generated, the results we see.' And also the action of setting a variable, X, to value x is simulated by replacing the structural equation for X with the equation $X = x$.

Finally, in general, $P(Y_x = y \mid x', y')$ cannot be expressed by a do-operator (i.e., expressed in the form $P(Y = y \mid do(X = x))$), but $P(Y = y \mid do(X = x))$ can be expressed as $P(Y_x = y)$. As Pearl, Glymour, and Jewell ([27], pp. 99-100) point out, $P(Y_{X=1} = y' \mid Z = 1)$ and $P(Y = y' \mid do(X = 1), Z = 1)$ are similar in form, but they characterize very different contents and answer very different questions.

4 Conclusion

I have presented a proof showing that $P^A(C)$ in CRT and $P^A(C)$ in CVPP have different extensions, so they are not equivalent. In turn, it is shown that CRT and CVPP are not equivalent as proved by Williams, so there is no such thing as CT.

We know that there are different versions of CT, for example, Santorio ([30]) also gives a version of the triviality result which, unlike Williams' triviality results, does not involve any specific way of cashing out suppositional credence for counterfactuals (ST and CRT), nor PP, as if the triviality results could be obtained from some weaker and less controversial assumptions. But Santorio actually deals with a more complex dimension of the counterfactual: the relationship between *would*-counterfactuals and *might*- counterfactuals. Indeed, because it does not involve the cashing out of counterfactual suppositions, the assumptions he presupposes are derived more from intuitive plausibility than from specific application contexts, and the refutation of Williams' CT does not constitute a solution to Santorio's CT.

But I think the point of this rebuttal to Williams' CT is that, within the framework of the SCM, the kind of "counterfactual" involved in CDT is at least algorithmically different from the paradigmatic counterfactual, and I am not convinced that this means that such counterfactuals are in fact indicative conditionals, but at least it shows that we have to be very careful when we prepared to equate some norms that apply to it with norms that apply to the paradigmatic counterfactual, as Williams' CT illustrates.

References

[1] E. W. Adams, 1965, "The logic of conditionals", *Inquiry*, **8(1–4)**: 166–197.

[2] E. W. Adams, 1975, *The Logic of Conditionals: An Application of Probability to Deductive Logic*, Dordrecht: Springer Science & Business Media.

[3] E. W. Adams, 1976, "Prior probabilities and counterfactual conditionals", *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, pp. 1–21, Dordrecht: Springer.

[4] R. Bradley, 2000, "A preservation condition for conditionals", *Analysis*, **60(3)**: 219–222.

[5] R. Briggs, 2017, "Two interpretations of the ramsey test", in H. Beebee, C. Hitchcock and H. Price (eds.), *Making a Difference: Essays on the Philosophy of Causation*, pp. 33–57, London: Oxford University Press.

[6] K. DeRose, 2010, "The conditionals of deliberation", *Mind*, **119(473)**: 1–42.

[7] D. Edgington, 2008, "Counterfactuals", *Proceedings of the Aristotelian Society* (*Hardback*), **108(1)**: 1–21.

[8] D. Edgington, 2011, "Conditionals, causation, and decision", *Analytic Philosophy*, **52(2)**: 75–87.

[9] A. Gibbard and W. L. Harper, 1978, "Counterfactuals and two kinds of expected utility", *IFS: Conditionals, Belief, Decision, Chance and Time*, pp. 153–190, Dordrecht: Springer.

[10] A. Hájek, 2014, "Probabilities of counterfactuals and counterfactual probabilities", *Journal of Applied Logic*, **12(3)**: 235–251.

[11] W. L. Harper and B. Skyrms, 1988, *Causation in Decision, Belief Change, and Statistics*, Dordrecht: Kluwer Academic Publishers.

[12] C. Hitchcock, 2013, "What is the 'cause' in causal decision theory?", *Erkenntnis*, **78(1)**: 129–146.

[13] C. Hitchcock, 2016, "Conditioning, intervening, and decision", *Synthese*, **193(4)**: 1157–1176.

[14] C. Hitchcock, 2022, "Causal Models", in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (*Spring 2022 Edition*), https://plato.stanford.edu/archives/spr2022/entries/causal-models/.

[15] J. M. Joyce, 1999, *The Foundations of Causal Decision Theory*, Cambridge: Cambridge University Press.

[16] D. Lewis, 1973, *Counterfactuals*, MA: Harvard University Press.

[17] D. Lewis, 1976, "Probabilities of conditionals and conditional probabilities", *The Philosophical Review*, **85**: 3.

[18] D. Lewis, 1979, "Counterfactual dependence and time's arrow", *Noûs*, **13(4)**: 455–476.

[19] D. Lewis, 1980, "A subjectivist's guide to objective chance", *IFS: Conditionals, Belief, Decision, Chance and Time*, pp. 267–297, Dordrecht: Springer.

[20] D. Lewis, 1981, "Causal decision theory", *Australasian Journal of Philosophy*, **59(1)**: 5–30.

[21] D. Lewis, 1986, "A subjectivist's guide to objective chance", in D. Lewis (ed.), *Philosophical Papers*, **Vol. 2**, pp. 83–113, London: Oxford University Press.

[22] D. Lewis, 1994, "Humean supervenience debugged", *Mind*, **103(412)**: 473–491.

[23] J. Liu, 2018, "How to escape triviality results?", *Studies in Logic*, **11(4)**: 56–82.

[24] C. Meek and C. Glymour, 1994, "Conditioning and intervening", *The British Journal for the Philosophy of Science*, **45(4)**: 1001–1021.

[25] J. Pearl, 2009, *Causality: Models, Reasoning, and Inference* (*2nd edition*), New York: Cambridge University Press.

[26] J. Pearl, 2013, "Structural counterfactuals: A brief introduction", *Cognitive Science*, **37(6)**: 977–985.

[27] J. Pearl, M.Glymour and N. P. Jewell, 2016, *Causal Inference in Statistics: A Primer*, New York: John Wiley & Sons.

[28] F. Ramsey, 1978, "Law and causality", *Philosophical Papers*, pp. 140–163, Cambridge: Cambridge University Press.

[29] R. Rooij, 2006, *Attitudes and Changing Contexts*, Dordrecht: Springer.

[30] P. Santorio, 2022, "General triviality for counterfactuals", *Analysis*, **82(2)**: 277–289.

[31] M. Schulz, 1975, *Counterfactuals and Probability*, London: Oxford University Press.

[32] W. Schwarz, 2014, "Proving the principal principle", in A. Wilson (ed.), *Chance and Temporal Asymmetry*, pp. 81–99, Oxford: Oxford University Press.

[33] W. Schwarz, 2018, "Subjunctive conditional probability", *Journal of Philosophical Logic*, **47(1)**: 47–66.

[34] T. Sider, 2010, *Logic for Philosophy*, New York: Oxford University Press.

[35] B. Skyrms, 1980, "The prior propensity account of subjunctive conditionals", *IFS: Conditionals, Belief, Decision, Chance and Time*, pp. 259–265, Dordrecht: Springer.

[36] R. Stalnaker, 1968, "A theory of conditionals", *IFS: Conditionals, Belief, Decision, Chance and Time*, pp. 41–55, Dordrecht: Springer.

[37] R. Stalnaker, 1970, "Probability and conditionals", *Philosophy of Science*, **37(1)**: 64–80.

[38] Q. Su, 2011, "On trivial results and constraints on conditionalization", *NTU Philosophical Review*, **41**: 113–133.

[39] J. Williams, 2012, "Counterfactual triviality: A Lewis-impossibility argument for counterfactuals", *Philosophy and Phenomenological Research*, **85(3)**: 648–670.

反事实的贫乏性与结构因果模型

吴小安

摘　要

威廉姆斯提出了反事实贫乏性问题。在本文中，我仔细考察了威廉姆斯的贫乏性结果所依赖的四个前提，并在结构因果模型的框架内，证明了这四个前提中的两个前提（CRT 和 CVPP）分别应用于两种不同类型的反事实条件句，所以在两个前提中所使用的 $P^A(C)$ 并不等价，从而证明了威廉姆斯的贫乏性结果并不成立。

吴小安　西北工业大学马克思主义学院
西北工业大学陕西省舆情信息研究中心
wuxiaoan1984@126.com