Studies in Logic, Vol. 15, No. 6 (2022): 49–65
PII: 1674-3202(2022)-06-0049-17

Causation as a Tool or Causation as a Target
——The Analysis of Pearl's and Lewis' Theory of Causation*

Zhanglyu Li Shangcheng Tang

Abstract. Judea Pearl's and David Lewis' theory of causation both hold important positions in the field of causation studies, but it seems their difference and applicability still need to be compared and discussed. Using Pearl's Logic of Structure-Based Counterfactuals, we analyze the three predicaments Lewis' theory of causation encountered, i.e. pre-emption, epiphenomena and cause-effect inversion, and show these problems can be answered easily with Pearl's theory. This "easy answer" reveals the logic preferences of the two theories: Pearl believes "the truth value doesn't influence the causal relation", but Lewis insists "the truth value do change the causal relation". Their logical preferences explain the major difference between Pearl's and Lewis' theory of causation: causation as a tool or causation as a target. Pearl's theory is more efficient to deal with the "tool-style" problems, which treats causation as a presupposed structure; while Lewis' theory is more suitable for "target-style" problems, which reject causation as an initial concept and try to find its metaphysical grounding.

1 Introduction

As human beings, we think with the help of causation, "some tens of thousands of years ago, humans began to realize that certain things cause other things and that tinkering with the former can change the latter."([9], p. 1) That is why pondering and interpreting the concept of "causation", became the main interest of so many philosophers, logicians and statisticians through the ages.

On the logic side of causation studies, David Lewis is a giant we could not neglect. In 1973, in his book *Counterfactuals* ([2]), Lewis used a logic system based on comparative similarity (now called "Similarity Logic" for short) to determine the truth value of counterfactual conditionals. Then Lewis interprets causation as causal dependence. In his opinion, causal dependence could be defined by counterfactual dependence, which could be analyzed in his logic system. Lewis' works still have a huge impact on the field of logic and philosophy studies.

21 years later, Lewis' challenger comes. As a computer scientist, Judea Pearl raised his theory of causation based on the Logic of Structure-Based Counterfactuals

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

Zhanglyu Li Institute of Logic and Intelligence, Southwest University
lizhanglv@126.com

Shangcheng Tang Institute of Logic and Intelligence, Southwest University
541103835@qq.com

*This work was supported by the National Social Science Fund of China (19BZX134).

(now called "Structure Logic" for short, [6]).[1] Pearl himself claimed that Lewis' Similarity Logic and his Structure Logic are identical for recursive systems. "In sum, for recursive models, the causal model framework does not add any restrictions to counterfactual statements beyond those imposed by Lewis' framework; the very general concept of closest worlds is sufficient."([8], p. 242) But Keith Markus thought there is a tricky point: this claim of logical equivalence has an implication that Pearl's theory as a successor superseded Lewis' theory. But the theory of causation is a broader concept than the logic embedded in it, so it's too early to reach the conclusion of "who-wins-out". "Indeed, Lewis' theory of causation …would still differ from Pearl's notion of causation even if they shared the same theory of counterfactual conditionals."([5], p. 444) We favor Markus's opinion, and think it's a fake question to discuss whether Pearl's theory superseded Lewis' or not. What interests us is the difference and applicability of the two theories. In this paper, after a summary of Lewis' and Pearl's theory of causation (Section 2 and half of Section 3), we will analyze and discuss: (a) For some problems Lewis' theory can hard to give a proper solution, can they be solved with Structure Logic? (the rest of Section 3) (b) From the different treatment of the two theories toward the same problem, can we dig out Pearl and Lewis' inner motive, i.e. their philosophical presupposition and even the different understanding of the concept of "causation"(Section 4). Based on our analysis and discussion, we will give a clarification about the difference and applicability of the two theories of causation. (Section 5)

2 Lewis' Theory of Causation and its Predicaments

2.1 Similarity logic and causation

Lewis' study of caution begins with the study of counterfactual conditionals. Some people may think Lewis took a detour, but through the analysis of counterfactual conditionals, he did seize the vital part of causation. In daily life, conditionals are our most common expression to convey the concept of causation. That is to say, if we think A and B have causal relation, we usually utter a conditional like "If A happens, then B will happen". This utterance not only set B after A temporally but also set B follows from A logically. For most of conditionals, if we want to prove or disprove the causal relation conveyed within them, we could use empirical facts to analyze whether they obey the logic rule "If the antecedent is true, then the consequent must be true"(Though there may always be new facts to disapprove them). But counterfactual conditionals cannot be proved or disproved by empirical facts, because what their antecedents expressed counter the fact: "If A happens (But A didn't

[1] To avoid confusion, a rephrase: Pearl's theory of causation based on Structure Logic and Lewis' theory of causation based on Similarity Logic. When we mention "Structure Logic" or "Similarity Logic" later, the theory based on it is also concerned.

happen in reality), then B will happen."[2] So, "how to determine the truth value of counterfactual conditionals", becomes the hardest part of causation studies. Lewis' Similarity Logic, exactly gives us a complete solution for the determination of the truth value of counterfactual conditionals.

To understand Similarity Logic, first we must understand Lewis' version of possible world semantics. Possible world semantics believes that our real world has some accessible worlds, and to determine the truth value of some particular propositions (propositions with modality, like $\Box\varphi$) on our world, we must consider the truth value of these proposition on the accessible worlds ($\Box\varphi$ is true on our world if and only if φ is true on all the accessible worlds of our world). Lewis also believes that if a world can be accessible from our world, this world must have some degree of similarity with our world.[3]

On the basis of possible world semantics, Lewis defined a counterfactual conditional operator, "$\Box\!\!\rightarrow$". Using this operator, we can formalize counterfactual conditionals and determine their truth value. For example, Bob's mother promised to buy him a new toy if he scored 90 on the Chinese test. After the result was announced, Bob only scored 89, so his mother didn't buy him a new toy. So, Bob said: "If I had scored 90 on the Chinese test, my mother would have bought me a new toy." This sentence is a counterfactual conditional. If we use φ to represent "Bob scored 90 on the Chinese test" and ψ to represent "Bob's mother bought him a new toy", then the sentence can be formalized as "$\varphi \mathrel{\Box\!\!\rightarrow} \psi$".

How to determine the truth value of "$\varphi \mathrel{\Box\!\!\rightarrow} \psi$"? Lewis uses the concept "comparative similarity". In his definition "$\varphi \mathrel{\Box\!\!\rightarrow} \psi$" is true at a world i (according to a given comparative similarity system) if and only if either: (1) no φ-world k belongs to S_i(the vacuous case), or (2) there is a φ-world k in S_i such that, for any world j, if $j \leqslant_i k$ then $\varphi \rightarrow \psi$ holds at j. ([2], p. 49) "φ-world" represents the world on which the formula φ is true, "S_i" represents the set of accessible worlds of world i, and"$j \leqslant_i k$" means that the distance between the world j and the world i on the concentric circle[4] is less than or equals to the distance between the world k and the world i (the closer it is to the world i, the more similar it is to the world i).

[2]Some people may think "A didn't happen" can prove the antecedent is false, then the whole counterfactual conditional is true by the rule of propositional logic. But this kind of treatment will lead us to paradoxes of implication: every conditional could be true, as long as it has a false antecedent.

[3]I take as primitive a relation of comparative over-all similarity among possible worlds. We may say that one world is closer to actuality than another if the first resembles our actual world more than the second does. ([1])

[4]In Lewis' graphical representation, the accessible worlds of our real world, are the points in a circle, and the center of the circle is the real world. Depending on the degree of similarity with the real world, countless concentric circles can be drawn, which form countless rings. The points in different rings have different similarity, and the closer the ring from the center, the more similar the points in it are to our real world.

In (1), there is no φ-worlds belonging to S_i, which is similar to the case where the antecedent is false in the substantive implication, when φ is false in all accessible worlds, regardless of the truth value of the ψ, $\varphi \boxdot\!\!\rightarrow \psi$ is true on the world i.

In (2), after we sort all the accessible worlds of the real world according to the degree of similarity with the real world, from largest to smallest, we could get a rank of worlds. As long as there is a φ-world which makes ψ true, and $\varphi \rightarrow \psi$ are true on all possible worlds between this world and the real world (including this world and the real world), we can conclude that the $\varphi \boxdot\!\!\rightarrow \psi$ is true on the real world. Using Bob's example, that is, in a certain possible world(which is an accessible world of our world), Bob scored 90 on the Chinese test (which makes this world become a φ-world), and his mother bought him a new toy (which means ψ is also true on this world), and on all worlds which have the same or more comparative similarity, the substantiative implication "$\varphi \rightarrow \psi$" are true, then the counterfactual conditional "$\varphi \boxdot\!\!\rightarrow \psi$" is true on our world.

As we can see, Lewis' use of comparative similarity is very intuitive. The hardest part in determining the truth value of a counterfactual conditional is that the situation described in the antecedent counters the situation of the real world. By interpreting the accessible world as "a world that has some degree of similarity to our real world", Lewis can stipulate the description of the antecedent be true or false on the accessible worlds, and successfully determine the truth value of counterfactual conditionals in the end.

After successfully determine the truth value of the counterfactual conditionals, Lewis uses Similarity Logic to interpret causation. First, he defines a relationship called "counterfactual dependence". "Let $A_1, A_2, \ldots$ be a family of possible propositions, no two of which are compossible; let $C_1, C_2, \ldots$ be another such family (of equal size). Then if all the counterfactuals $A_1 \boxdot\!\!\rightarrow C_1, A_2 \boxdot\!\!\rightarrow C_2, \ldots$ between corresponding propositions in the two families are true, we shall say that the C's depend counterfactually on the A's."([1], p. 561) Then, Lewis defines Causal dependence. "If a family $C_1, C_2, \ldots$ depends counterfactually on a family $A_1, A_2, \ldots$ in the sense just explained, we will ordinarily be willing to speak also of causal dependence."([1], p. 561) For example, assuming the possible indoor temperature is 18 degrees to 26 degrees, then every possible temperature can be a proposition, which we call a family of propositions $A_1, A_2, \ldots$. At the same time, every reading displayed on the thermometer from 18 degrees to 26 degrees can also be a proposition, which we call a family of propositions $C_1, C_2, \ldots$. The C's counterfactually depend on A's, which is the thermometer readings counterfactually depend on indoor temperature.[5] Or we can

[5]Suppose the temperature indoors is 20 degrees, and a person in the room said to himself: "If the indoor temperature was 25 degrees just now, the thermometer would show the number 25 (the reading of the thermometer would be 25 degrees)". This is a counterfactual conditional, and it does express a causal relation between the room temperature and the thermometer readings.

directly say: the indoor temperature is the cause, the thermometer reading is the effect.

Thus, based on Lewis' analysis of causation, we can conclude: the causation is defined as a causal dependence, and the causal dependence is defined as a counterfactual dependence between a family of propositions $(A_1, A_2, \dots)$ and the other family of propositions $(C_1, C_2, \dots)$. To avoid confusion, we need to make an explanation. If the event A and the event C only have two situations (occur/doesn't occur), then the family of propositions of event A only contains two propositions: A and $\neg A$, and the family of propositions of event C only contains C and $\neg C$. Suppose the event A is the cause of the event C, according to our definition above, C's must counterfactually depend on A's. Then only two formulas need to be true, which are $A \boxdot\!\!\rightarrow C$ and $\neg A \boxdot\!\!\rightarrow \neg C$. [6]

2.2 The three predicaments Lewis' theory encountered

Lewis' definition of causation is both mathematically precise and philosophically novel. But while receiving applause, Lewis' theory of causation encountered at least the following three predicaments. ([10])

First, Lewis' theory of causation is difficult to explain the pre-emption problem. The pre-emption problem is that both events can cause the same result, but the occurrence of one event prevents the occurrence of the other. For example, A and B are playing with the ball, and when the ball flies out of the field, A and B both run to pick up the ball, B sees that A runs faster, so he stops running, and lets A pick up the ball and bring it back to the field. If the effect is "the ball goes back to the field (C)", then "A picks up the ball (A)" is the cause. The predicament Lewis' theory encountered is that A and C only satisfy $A \boxdot\!\!\rightarrow C$,[7] but not $\neg A \boxdot\!\!\rightarrow \neg C$ ($\neg A \boxdot\!\!\rightarrow \neg C$ is false on our world (the world i) because when there is an accessible "$\neg A$-world" k, $\neg C$ could be false on k, which violated condition (1) and (2). In other words, even if A didn't pick up the ball, the ball will also go back to the field because of B's act.), thus not meeting the definition of causal dependence. Lewis' solution is to define a new concept "influence", and thinks under normal circumstances, causation could be defined as causal dependence, but for pre-emption, causation is defined as "influence". "let us say that A influences C if and only if there is a substantial range $A_1, A_2, \dots$ of different not-too-distant alterations[8] of A (including the actual alteration of A) and

[6] Here we simplified Lewis' symbol, Lewis uses e to represent an event, and $O(e)$ to represent the proposition corresponding to the event.

[7] $A \boxdot\!\!\rightarrow C$ is true on our world (the world i) because when i is an "A-world", C must be true on i, and there is no world j which satisfied $j \leqslant_i i$, then the condition (2) is met. Since A is true in the real world, $A \boxdot\!\!\rightarrow C$ doesn't seem to be counterfactual conditional according to our definition. However, Lewis believes this is also a special kind of counterfactual conditionals, which are "counterfactual conditionals that are not counterfactual". ([2])

[8] The alteration (of an event) is a version of the event or an alternative to the event.

there is a range $C_1, C_2, \ldots$ of alterations of C, at least some of which differ, such that if A_1 had occurred, C_1 would have occurred, and if A_2 had occurred, C_2 would have occurred, and so on."([4], p. 190)[9] Taking the above "pick up the ball" problem as an example, A influences C, that is, when A occurs, C will also occur due to the influence of A, and the way it occurs is "A picks up the ball and brings the ball back to the field"; When A does not occur, C will also occur, but due to the influence of A, the way it occurs becomes "A didn't pick up the ball, but B picks up the ball and bring the ball back to the field". Lewis' definition of influence is vague. Because the "alteration" is a variable that is hard to characterize and exhaust, like the measure of "different not-too-distant" in Lewis' definition, which is actually different in different people's minds. And by its definition, in some cases, the change of A will cause the change of C. But in other cases, the change of A won't change C. We can't exhaust all these situations in practice, for example, "A kicked the ball when he went to pick up the ball" is an alteration of "A went to pick up the ball", then the corresponding result may be "the ball was kicked far away and did not go back to the field", or "the ball was kicked against the wall and bounced back to the field", both of which are alterations of "the ball goes back to the field", but which result should the cause be corresponded to? To answer this question, we must check the details like the strength and the angle of the kick, the distance between the wall and the field, and so on, which makes more and more alterations of "A went to pick up the ball". Therefore, "influence" is only an intuitive concept, but cannot be clearly characterized.

Second, Lewis' theory of causation lacks a refined answer for the epiphenomena problem: when A is a necessary and sufficient condition for C, if A is also the cause of D, then Lewis' theory will draw the wrong conclusion that "C is the cause of D". The wrong conclusion arises because, when C is true on any accessible world k of our world, we can infer that A is also true on the world k, so D is true on k, too (because $A \mathrel{\Box\!\!\rightarrow} D$ is true on our world, which determined D must be true on every accessible "A-world"); when$\neg C$ is true on any accessible world k of our world, we can infer that $\neg A$ is also true on world k, so $\neg D$ is true on k, too (because $\neg A \mathrel{\Box\!\!\rightarrow} \neg D$, which determined $\neg D$ must be true on every accessible "$\neg A$-world"). Then the two formulas $C \mathrel{\Box\!\!\rightarrow} D$ and $\neg C \mathrel{\Box\!\!\rightarrow} \neg D$ are true on our world, which satisfied Lewis' definition of counterfactual dependence, so we could get the conclusion that C is the cause of D. But in many cases, C is not the cause of D. Suppose that"the positive and negative charges in the thunder cloud touch" (A) is a necessary and sufficient condition for "we hear thunder" (C), and is the cause of "we see the flash" (D),[10] but C and D have no

[9]For simplicity, we changed the "C" in Lewis' original text to "A" and the "E" to "C".

[10]In most of the time, when the positive and negative charges in the thunder cloud touch, people would not see the thunderbolt directly but see the flash in the sky. To clarify this is because "the positive and negative charges in the thunder cloud touch" is the necessary and sufficient condition for "people see the thunderbolt", but is only the sufficient but not necessary condition for "people see the flash".

causal relationship. Lewis' theory will draw the wrong conclusion that "We see the flash because we heard the thunder". Lewis' treatment of this problem is denying that D causally depends on C, that is, although his theory gives out the rule that "when counterfactual dependence is established, the causation is established", the rule can be broken when it is seriously violated the reality. So, we could deny that C is the cause of D even if $C \ \Box\!\!\rightarrow \ D$ and $\neg C \ \Box\!\!\rightarrow \ \neg D$ both are true. This treatment is what we called "rule-breaking" treatment. In Lewis' view, violating the rules and allowing fewer "miracles" to occur is closer to reality than obeying the rules and allowing more miracles to occur. This "rule-breaking" treatment is quite crude.

Third, Lewis' theory of causation will lead to "cause-effect inversion" between a necessary and sufficient condition and its result. When C causally depends on A, if we suppose that A is a necessary and sufficient condition for C, then when C is true on any accessible world k of our world (the world i), we can infer A is also true on k; we can also infer that $\neg A$ is true on k when $\neg C$ is true on k. Then we have the two formulas $C \ \Box\!\!\rightarrow \ A$, $\neg C \ \Box\!\!\rightarrow \ \neg A$ are true on world i, which inverses the cause and effect, making C as the cause of A. However, most of time, this kind of inversion is wrong in the real world. For example, "the temperature reaches above 0 degrees" is a necessary and sufficient condition for "the ice cube melts", but "the ice cube melts" is obviously not the cause for "the temperature reaches above 0 degrees". Lewis' treatment is to deny A causally dependent on C , that is, the "rule-breaking" treatment we mentioned above.

Next, we will use Structure Logic to analyze these three predicaments.

3 Pearl's Theory of Causation and its Answer to Lewis' Predicaments

According to Pearl himself, in his early years of research, he hoped to explain causation with the help of probability, but it didn't work. "Today, my view is quite different. I now take causal relationships to be the fundamental building blocks both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world."([8], p. xvi) Like Lewis, Pearl used counterfactual conditionals as a breakthrough point in constructing his theory of causation.[11]

Because we could also see the flash in the sky under the condition like "shooting a flare bomb", "turn on the spotlight for a second".

[11] "How do scientists predict the outcome of one experiment from the results of other experiments run under totally different conditions? Such predictions require us to envision what the world would be like under various hypothetical changes and so invoke counterfactual inference. Though basic to scientific thought, counterfactual inference cannot easily be formalized in the standard languages of logic, algebraic equations, or probability."([8], p. 202)

3.1 Structure Logic and causation

Intuitively, Pearl's theory of causation hopes to give a model at first, which characterizes the causal relations that already exist, then modify the cause event of the model according to the state of affairs described in the antecedent of the counterfactual conditional, and finally, check whether the modified model has a causal relation that makes the counterfactual conditional true. Pearl's theory can be expressed in two ways: the causal diagram and the symbolic language.

The causal diagram is also named the directed acyclic graph (*DAG*), which gives a graphical representation of causal relations. "The causal diagrams are simply dot-and-arrow pictures that summarize our existing scientific knowledge. The dots represent quantities of interest, called 'variables', and the arrows represent known or suspected causal relationships between those variables—namely, which variable 'listens' to which other."([9], p. 7) When we have a "causal intuition" about the relation between events, that is, we suspect that there is a causal relation between events, we can characterize our causal intuition to causal diagrams, and correspond the related data we gathered to the diagrams. For example, we think that the fire produces the smoke, and the smoke causes the alarm to ring. This causal intuition contains three variables: "fire", "smoke", "alarm". Then we can draw a causal diagram like "fire→smoke→alarm", and correspond with the data like "fire = 1 (that is, fire occurs), smoke = 1, alarm = 1", "fire = 0, smoke = 0, alarm = 0".

The symbolic language is the Structural Logic that gives a mathematical characterization about causation. Pearl's causal model is $M = \langle U, V, F\rangle$. In this model, U, V are sets of variables, which could be understood as events in the causal relation. The difference between U and V is that U is the set of background variables, and the variables in it are determined by factors outside the model. While V is a set of endogenous, which are determined by variables in the model. F is the set of functions that determine causal relations. Having set the model M, the causal relations we concerned are determined. But to check the truth value of the counterfactual conditional, we must modify our model, i.e., intervene the variable. So, Pearl introduced the concept of "submodel". A submodel is a model which replaced the related event in the original model with the antecedent of the counterfactual conditional. The process of intervention is like this: first, we pick a set X (as the event we want to replace), which is the subset of V. Then, define the submodel $M_x = \langle U, V, F_x\rangle$, where the F_x is a modification of F—which deletes all functions mapped to variables in X, and maps all variables in X to x, that is, control variable X to a fixed value x. The modification of F can be expressed as $F_x = \{f_i : V_i \notin X\} \cup \{X = x\}$. This action of intervention is expressed as $do(X = x)$. Finally, the truth value of the counterfactual conditional is determined by checking whether the consequent is true in the submodel. Let Y be the set of variables we want to check (Y is a subset of V), $Y_x(u)$ is the solution for (the event) Y which can be obtained from the set of functions F_x, we can also say

that $Y_x(u)$ is the potential response of Y to the action $do(X = x)$. How to check the result? Pearl puts like this: "Let X and Y be two subsets of variables in V. The counterfactual sentence 'Y would be y (in situation u), had X been x' is interpreted as the equality $Y_x(u) = y$, with $Y_x(u)$ being the potential response of Y to $X = x$."([8], p. 204) In other words, if $Y_x(u) = y$ in the submodel, then the counterfactual is true.

In the next section, we will analyze whether Structural Logic can have a better answer for the three predicaments encountered by Lewis' theory.

3.2 Answer the problem of "pre-emption" with Structure Logic

In the "pick up the ball" case we mentioned before, set "A picks up the ball" as A, "B picks up the ball" as B, and "the ball goes back into the field" as C. In fact, A and B also have a common cause, that is, "the ball flies out of the field", set it as Q. We could draw a causal diagram like Figure 1:

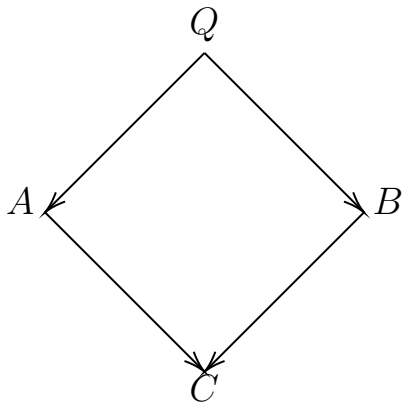

Figure 1

And our causal model M:[12]

$$\begin{array}{ll} & (Q) \\ A = Q & (A) \\ B = Q & (B) \\ C = A \vee B & (C) \end{array}$$

The predicament Lewis encountered is the causal dependence between C and A can't be established. For Structure Logic, to determine the causal relation between A and C, we need to check "If A picked up the ball, and B didn't pick up the ball, would

[12]"$A = Q \quad (A)$" means A is determined by Q. The reason for Pearl using "$=$" but not "$\Leftarrow$" is the equal sign shows that we could do abduction easily, "(A)" means A is the one that is determined by the other.

the ball go back to the field?", which lead to the intervention of B. After $do(B = 0)$, by definition, we need to delete the equation "$B = Q \quad (B)$"(the arrow from Q to B), and fix $B = 0$ (which means $\neg B$ is true). Then we get Figure 2:

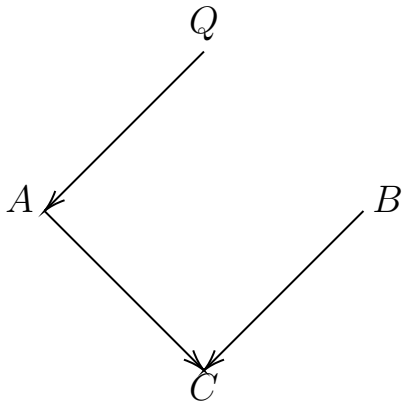

Figure 2

And the submodel $M_{\neg B}$:

$$\begin{array}{ll} & (Q) \\ A = Q & (A) \\ \neg B & (B) \\ C = A \vee B & (C) \end{array}$$

After the intervention, we need to check whether C is true when A is true. From "$C = A \vee B \quad (C)$" and "A is true (our condition)", we can infer that "C is true", which proves that A is the reason for C.

3.3 Answer the problem of "epiphenomena" with Structure Logic

In the "thunder cloud" case we mentioned before, A is the necessary and sufficient condition for C, and A is the cause of D. But D also has other causes, like "shooting a flare bomb (B)". We could draw a causal diagram like Figure 3:

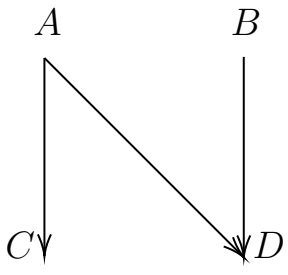

Figure 3

And our causal model M:

$$\begin{array}{ll} & (A) \\ & (B) \\ C = A & (C) \\ D = A \vee B & (D) \end{array}$$

In this problem, Lewis' predicament is: with the condition and Similarity Logic, we could draw the wrong conclusion that D is causally dependent on C.

In Structure Logic, Figure 3 has a "fork junction", which shows A as the common cause of C and D. We could also say A is a confounder of C and D. Confounders can cause two unrelated variables to have a spurious correlation. In Figure 3, A as a confounder, creates a spurious correlation between C and D, that is, "the positive and negative charges in the thunder cloud had touch" as the confounder, makes the unrelated "we hear thunder" and "we see the flash" have a spurious correlation. In Lewis' theory, this spurious correlation lets "we hear thunder" become the reason for "we see the flash".[13] In Pearl's opinion, to eliminate this spurious correlation, we need to deconfound the confounding factor. The deconfounding method is to close the back-door path between C and D. For Pearl, "A set of variables S is said to satisfy the back-door criterion relative to an ordered pair of variables (X_i, X_j) in a *DAG* if: (1) No node in S is a descendant of X_i, and (2) S blocks every path between X_i and X_j which contains an arrow into X_i."([7], p. 679) Or a simpler definition: "A back-door path is any path from X to Y that starts with an arrow pointing into X."([9], p. 158) In Figure 3, to eliminate this spurious correlation of C and D, we need to find the back-door path, which is "$C \leftarrow A \rightarrow D$", the only back-door path between C and D. And conditioning on the variable A to a definite value can successfully close this path and eliminate spurious correlations. We have two assignments of A's value: (1) $A = 1$, then, from "$C = A \quad (C)$" and "$D = A \vee B \quad (D)$", we can infer that $C = 1$ and $D = 1$, that is, when the positive and negative charges in the thunder cloud touch, we will hear thunder and we will also see the flash; (2) $A = 0$, from "$C = A \quad (C)$", $C = 0$ can be inferred, but we could not infer the value of D, because "$D = A \vee B \quad (D)$" and we don't know the value of B. And if $B = 1$ at this time, we would get $D = 1$, that is, when the positive and negative charges in the thunder cloud didn't touch, we would not hear thunder, but we would see the flash, because a flare bomb had shot into the sky. So, when we conditioning on the variable to $A = 0$, the spurious correlation between the two variables C and D disappears. By deconfounding, the causal relationship in Figure 3 is defended, i.e., A is a common cause of C and D, but there is no causal relation between C and D.

[13] As we mentioned before, though with Similarity Logic we can infer this causal relation, Lewis himself denied it. His justification is this kind of "miracle" is not allowed in our real world, so confront this certain case we should break the rule (of his theory of causation). We do think this justification makes sense, but not satisfied for its crudity.

3.4 Answer the problem of "Cause-effect inversion" with structure logic

The Problem of "Cause-effect inversion" is a subproblem of the problem of "epiphenomena". Since Lewis defined causation as causal dependence, when A is the necessary and sufficient condition for C, based on his rules, C causally depends on A. But with his rules we could also infer that A causally depends on C, then the cause-effect inversion occurs.

But for Structure Logic, as we have analyzed in the problem of epiphenomena, the causal diagram is a presupposed structure. When we set "$C = A \quad (C)$", even if the necessary and sufficient condition let us could infer the truth value of the cause (A) from the truth value of the effect (C), the equation would not be changed, and still provide A is the cause of C, but C isn't the cause of A. Thus, for Structure Logic, the problem of "Cause-effect inversion" is not a predicament at all.

But the problem of "Cause-effect inversion" actually reveals the difference between Structure Logic and Similarity Logic. For Similarity Logic, the truth value could influence the causal relation. Because the causal relationship is defined by Lewis as counterfactual dependence, which is determined by the truth value of the variables on the accessible worlds, so different truth value assignments can produce different causal relations. But for Structure Logic, the truth value does not influence the causal relation. Because the causal diagram as a presupposed structure, would not change by different truth values. We would have a detailed discussion in Section 4 about where this logical difference could lead us, and we think the answer is the major difference between Pearl's and Lewis' theory of causation: causation as a tool or causation as a target.

4 The Major Difference between Pearl's and Lewis' Theory of Causation

As we have shown, the predicaments Lewis' theory encountered can be cleared up easily with Pearl's Structure Logic. The reason is Pearl's causal relations are preset, stable functions, which won't be influenced by truth value of variables. But if the problems can be solved so easily, why Lewis didn't choose a preset causal structure like Pearl but insist on his complex, "truth-determine-causation" way? To answer this question, we need to dig out the inner motive of Lewis (and Pearl). This route leads us to their grounding theory of causation, and their different understanding of the concept of "causation" are exactly the major difference of the two theories of causation.

4.1 Pearl's theory: causation as a tool

The precursor of Pearl's theory of causation is "path analysis" invented by Sewall Wright. The main attack "path analysis" received is: can we get an objective

conclusion from a subjective, presupposed causal structure? Wright's justification is not absolute, he just claims: "We can use the diagram in exploratory mode; we can postulate certain causal relationships and work out the predicted correlations between variables. If these contradict the data, then we have evidence that the relationships we assumed were false."([9], p. 79) But how that causal diagram comes up in the first place? Wright gives no answer. To support Wright, Pearl comments: "Wright understood from the very beginning that causal discovery was much more difficult and perhaps impossible."([9], p. 80) In other words, Wright is satisfied with the causal diagram as a useful black box. Facing the same attack, Pearl also gives his own justification: "The very fact that people communicate with counterfactuals already suggests that they share a similarity measure, that this measure is encoded parsimoniously in the mind, and hence that it must be highly structured."([8], p. 239) The "measure" in his words is a causal diagram. This may answer "why the diagram is useful" (because we share the same measure), but still didn't answer "how the measure (causal diagram) comes up". So, Pearl's theory of causation didn't give us a full grounding theory of causation, he may give a slight thought about the origin and composition of causation, but most of his effort was put to leverage causation as a tool. To make it more specific, the reason Pearl raised his Structure Logic and theory of causation, is that he despises most statisticians only care about "correlation" but neglect the "causal intuition" in our mind. Structure Logic could characterize the "causal intuition" as *DAG*, which could not only compute "correlation", but also "climb the ladder of causation" – compute "intervention" and "counterfactuals". Pearl believes by using "causation" in this way, we could get more knowledge about this world. But how that "causal intuition" emerges? This is not the interest of Pearl's theory. Avoiding this question, and using a presupposed causal structure from the beginning of the research, let Structure Logic have the feature "truth value doesn't influence the causal relation", which makes the predicaments Lewis' theory encountered could be solved easily in Structure Logic.

Since the causal relation (causal diagram) is presupposed before the research, Pearl' s "theory of causation" is more like an abstract title, at least the grounding theory of causation is not included. If we really want Pearl to give us an answer about the origin and composition of causation, though he used "causal intuition" in his works, he may still tend to think causation is a mind-irrelated, objective entity. "counterfactuals are not based on an abstract notion of similarity among hypothetical worlds; instead, they rest directly on the mechanisms (or 'laws', to be fancy) that produce those worlds and on the invariant properties of those mechanisms."([8], p. 239) "The formalization of counterfactual inference requires a language within which the invariant relationships in the world are distinguished from transitory relationships that represent one's beliefs about the world."([8], p. 202) To summarize, Pearl's interest is not the origin and composition of causation, but using causation as a tool.

4.2 Lewis' theory: causation as a target

Opposite to Pearl, the usage of causation is just not Lewis' interest. For Lewis, a presupposed, reality-based causal model like Pearl's is not enough to fulfill his philosophical pursuit. Instead, Lewis wants to discuss the truth value of a conditional under various states of affairs, i.e. the truth value of a conditional on many possible worlds. We can even make an analogy like this: in Lewis' model, the state of affairs on every accessible world, can correspond to a Pearl's model. The reason Lewis choose a detour to define the causal relation as counterfactual dependence, exactly lies in that counterfactual dependence involves various state of affairs, thus "causation" can perfectly ground on his metaphysics: through the path of "supervenience", he wishes to ground all the proposition about worlds on the "mosaic"([3]). "Mosaic" is the arrangements of particles(or in his more abstract notion, arrangements of points/qualities) in this world. In a specific moment, the arrangement of particles in the world is definite, which means every particle takes a specific space. So, all the changes in the world are just different arrangements of particles. These particles, like a mosaic, composed the world we could perceive. To relate counterfactual conditionals with the mosaic of the world, Lewis gives his grounding theory of causation:

First, in a specific moment, the mosaic of the world is definite. So, the truth value of every proposition about the world at this moment would be determined. For instance, the proposition "Subsolar point is directly on the Equator" would be determined "true" at the spring equinox, because at this moment the particles composed of the sun, the earth and other parts of the universe take a specific space, let the mosaic of the world shows the fact that subsolar point is directly on the Equator.

Second, when the arrangement of particles changes, the mosaic of the world would also change, which changes the truth value of the propositions of the world. The proposition "Subsolar point is directly on the Equator" would be determined "false" at June solstice, because the arrangement of particles is different from the spring equinox.

Third, causation is uttered by us in the form of the conditional, and the truth value of the conditional is determined by the states of affairs. So, the truth value of the conditional is determined by the mosaic of the world.

In the end, the different mosaics of the world correspond to different possible worlds. To prove a causal relation expressed by a conditional is to determine the counterfactual dependence between the antecedent and the consequent, which need to analyze the truth value of the antecedent and the consequent on different worlds. When the counterfactual dependence is determined, the causal relation is proved. This process answers the origin and composition of causation: causation comes from counterfactual dependence, and counterfactual dependence comes from the mosaic of the world. So, the causation is essentially grounded on the mosaic of the world.

So, in Lewis' eyes, causation is a target. He wants to break up the complex

"caution" to fundamental facts of the world. To fulfill this goal, he must rely on counterfactual dependence and Similarity Logic. Even in danger of those predicaments, he would just disobey the rules, but not discard his whole theory. Deep down in his theory of causation, is the giant iceberg of his "mosaic" metaphysics, that's what we need to see.

5 Conclusion

"Causation as a tool or Causation as a target", is the major difference between Pearl's and Lewis' theory of causation we summarize. Understanding "causation" as a tool, let Pearl's theory has rich practical value, which transformed classical statistics into "causation-friendly" version, and helped us conclude more scientific result; Understanding "causation" as a target, let Lewis' theory offer us a clear philosophical interpretation of the concept causation, which grounding the causation on the metaphysical basis of "mosaic".

This major difference, when looking at the side of logic, is Pearl's and Lewis' different answer to the question "Do truth value influence the causal relation or not?" Understanding "causation" as a tool, let Pearl presuppose the causal structure before the research, which embedded the causal relation as stable functions. This treatment of causation let him believe "The truth value doesn't influence the causal relation", which makes the predicaments Lewis' theory encountered can all be answered easily. Understanding "causation" as a target, let Lewis define causal relation as counterfactual dependence, which is composed by a set of propositions. After supervene the counterfactual dependence on the mosaic of the world, "The truth value does change the causal relation" becomes the rule he must obey. So, he is inevitable to face the predicaments like pre-emption, epiphenomena and cause-effect inversion.

From the major difference to the difference on the logic side, then to the different treatment of the specific problems, this route of analysis just proved the respective applicability of the two theories, we summarize as follows: when we have a question related to causation, we could first distinguish it as a "tool-style" question or a "target-style" question. The "tool-style" question is questions like "Have known the cause, how to predict the effect" "Have known the effect, how to find the causal chain" "When the cause didn't happen, the effect would be changed or not", which recognizing causation has already existed, and what we need to do is get more knowledge with the help of causation. Pearl's theory is more suitable for these "tool-style" questions. The "target-style" question is questions like "What basis does the causation ground on" "Is the causation a mind-related entity, or an objective structure" "How could people understand the notion of causation", which don't treat causation as a ready-to-use concept, but think causation need more fundamental explanation. Lewis' theory is more suitable for these "target-style" questions.

References

[1] D. Lewis, 1973, "Causation", *The Journal of Philosophy*, **70(17)**: 556–567.

[2] D. Lewis, 1973, *Counterfactuals*, Malden: Blackwell.

[3] D. Lewis, 1986, *Philosophical Papers*, New York, Oxford: Oxford University Press.

[4] D. Lewis, 2000, "Causation as influence", *The Journal of Philosophy*, **97(4)**: 182–197.

[5] K. A. Markus, 2021, "Causal effects and counterfactual conditionals: Contrasting Rubin, Lewis and Pearl", *Economics & Philosophy*, **37(3)**: 441–461.

[6] J. Pearl, 1994, "From bayesian networks to causal networks", in A. Gammerman (ed.), *Bayesian Networks and Probabilistic Reasoning*, pp. 157–182, London: Alfred Walter Ltd.

[7] J. Pearl, 1995, "Causal diagrams for empirical research", *Biometrika*, **82(4)**: 669–688.

[8] J. Pearl, 2009, *Causality: Models, Reasoning, and Inference*, New York: Cambridge University Press.

[9] J. Pearl and D. Mackenzie, 2018, *The Book of Why: The New Science of Cause and Effect*, New York: Basic books.

[10] W. Zhang 张文琴, 2018, *Counterfactuals and David K. Lewis's Philosophy of Logic* 大卫·刘易斯逻辑哲学思想研究——以反事实条件句为中心的考察, Shanghai: Shanghai Academy of Social Sciences Press.

以因果为工具与以因果为对象
——珀尔与刘易斯的因果理论之辨

李章吕　　唐上程

摘　要

珀尔的因果理论与刘易斯的因果理论在当今因果关系研究领域都具有重要地位，但学界对两种理论的差异与适用范围的讨论是不足的。通过使用珀尔的结构因果逻辑处理刘易斯因果理论遭遇的三大疑难，即“抢先”问题、“副现象”问题、“充分必要条件下倒果为因”问题，发现上述疑难都能被轻易消解。“轻易消解”的深层次原因是：在逻辑层面，珀尔认为“真值不影响因果关系”，而刘易斯认为“真值会影响因果关系”。这种逻辑层面的差异解释了珀尔和刘易斯因果理论的最大差异：以因果为工具与以因果为对象。珀尔的因果理论适合处理“因果工具型”问题，该类问题的特点在于默认了因果关系的存在；刘易斯的因果理论适合处理“因果对象型”问题，该类问题的特点在于不认为“因果”是一个可以不加分析而被使用的初始概念。

李章吕　西南大学逻辑与智能研究中心
lizhanglv@126.com
唐上程　西南大学逻辑与智能研究中心
541103835@qq.com