A new theory of causation based on probability distribution determinism

Chong Liu*

Abstract

The concept of causation is essential for understanding relationships among various phenomena, yet its fundamental nature and the criteria for establishing it continue to be debated. This paper presents a new theory of causation through a quasi-axiomatic approach. The core of this framework is *Probability Distribution Determinism* (PDD), which updates traditional determinism by representing states of affairs as probability distributions, with the '*if... then...*' function serving as its foundational definition. Based on PDD, by merely using appropriate naming strategies, it is possible to derive systems in which the structural characteristics of relationships among things closely resemble those in the real world, such as having various forms of nested hierarchies. Additionally, there are two related yet distinctly different contexts about relationships in PDD: one emphasizes the potential influence of conditions on outcomes in the general sense, while the other focuses on attributing responsibility for the state changes in specific scenarios. The formula for determining the relationship in the general sense is established as $S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$. Subsequently, within the PDD framework, the paper clarifies the legitimate use of a series of concepts related to causation in those two contexts, thus encompassing the entire detailed connotation of the concept of causation. The comparison with other theories of causation and the

*Email: john.liuchong@gmail.com

analysis of cases of application demonstrate that the new theory applies not only to situations where other theories are competent but also to situations where they are not. This suggests that, although certain aspects within the new framework may require further analysis, it provides a highly promising direction for a deeper understanding of causation.

Keywords: causation, probability distribution, determinism.

1 Introduction

The concept of causation is fundamental to understanding and interpreting the relations among things. However, there remains widespread debate about what exactly the connotation of causation is and how to determine whether there is a causal connection among things (Cartwright, 2004).

This paper proposes a new theory of causation by employing a quasi-axiomatic approach. Traditional theories of causation often rely on empirical induction or linguistic conceptual analysis. However, the limitations of human experience can restrict our perspective in understanding issues (Wittgenstein, 2009; Kuhn, 2012; Silver et al., 2017). In contrast, a quasi-axiomatic approach will likely overcome biases and analytical limitations imposed by human experience and natural language, resulting in a theory with broader explanatory power.

Although this new theory has drawn a wealth of insights from other theories of causation and evolved through critical reflection on them, its internal logic is independent. Therefore, the discourse structure in this paper primarily adheres to the inherent logic of the theory's establishment, while comparative discussions with other theories will be positioned as secondary, woven throughout the various sections.

Section 2 analyzes the limitations of traditional views that consider specific events or probabilities of events as the subjects of causal statements, pointing out that the state of

affairs can be expressed as a *probability distribution* over all possible specific values.

Building on this foundation, Section 3 revises traditional determinism into *Probability Distribution Determinism* (PDD), within which the outcomes may be either non-probabilistic or probabilistic. Subsequently, the *laws of nature* can be understood as a special mode of PDD, and Hume's *regularity theory* of causation is refuted. Furthermore, *free will* is discussed and shown to be compatible with PDD.

On the basis of PDD, Section 4 uses only naming strategies to construct multiple systems with structures of relationships similar to those in the real world, both of which have nested hierarchies. This similarity demonstrates the rationality of the assumption that PDD reflects the real world. Furthermore, we get insights for analyzing a series of issues about causation, such as the *additivity* of causal effects, *transitivity* and *ontological explanations* of causal relationships, and whether *temporal property* is necessary for causation. In addition, a critical analysis is conducted on the value and limitations of *mechanism* and *causal structural equations*.

Section 5 analyzes *two different contexts* for statements regarding relationships in PDD: one emphasizes the potential influence of conditions on outcomes in the general sense, while the other focuses on attributing responsibility for the state changes in specific scenarios. Then, it explains why *counterfactual theories of causation* failed.

The central theme from Section 2 to Section 5 is the metaphysical analysis of PDD. Based on this, Section 6 undertakes an analysis from the epistemological perspective for the relationship between things in the general sense in PDD, and gets the formula for determining it as $S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$. Eventually, Section 7 explicates the proper use of a series of concepts related to causation and completes the system of this new theory of causation.

In order to further illustrate the explanatory power of the new theory, Section 8 compares it with some highly influential theories of causation, such as *Randomized Controlled Trials* and *Interventionism*. Section 9 concisely demonstrates how the new theory can be applied

in scientific research where those theories cannot.

2 State of affairs: expressed as probability distribution

In the formation and evolution of concepts in natural language, the associated experiential contexts are diverse and dynamic, leading to the phenomenon where words in natural language often carry multiple meanings. This ambiguity is particularly evident in the concept of causation, one of the most commonly used concepts in daily life, as pointed out by Russell (1913) and Cartwright (2004). Thus, it may ultimately be in vain to provide a unified and precise definition only based on analyzing the meaning of causation in scattered experiential contexts. (Wittgenstein, 2009).

However, existing theories of causation have been developed more or less in this manner. For instance, for these theories, what exactly are the *relata* of causation appears to be a fundamental starting point question that is often positioned at the initial stages of their analysis. While some traditional theories posited causal relations are connections between substances such as individual objects or events (Mumford and Anjum, 2013), contemporary scholars argue that these explanations are not applicable to relationships between abstract properties, and then consider the primary relata of causation as *variables*, which are symbols that represent entities, objects, events, or properties (Cartwright, 1979; Hitchcock, 1993; Woodward, 2003, p. 17; Pearl and Mackenzie, 2018, p. 7).

Although treating variables as the relata of causation has shown good applicability in scientific practice, the effectiveness is primarily empirically inductive. Whether this approach can generalize all cases lacks rigorous demonstration, implying the potential for exceptional cases it cannot handle. What is more, the accuracy of the definition of the relata fundamentally impacts the explanatory power of the theory of causation based on it. To avoid basing the final theory on unstable foundations, this paper proposes temporarily suspending the determination of what exactly the relata of causation are at the beginning of the analysis,

similar to temporarily suspending the definition of causation, as the two questions are closely related.

In contrast, another question, similar yet fundamentally different to the question about the relata of causation, remains meaningful: If there is some sort of connection between things rather than being independent of each other, what words can we use appropriately to describe everything related to that relationship, and how to identify it? And this is exactly what we are concerned with and will attempt to address in this article.

This paper's analysis begins by hypothesizing an observer without background knowledge or presuppositions. For such an observer, unaware of any connections or influences between things, the most fundamental information is the *state of affairs* at each moment. Therefore, the paper first analyzes what a complete and sufficient scientific expression of the state of affairs is.

For a certain state of affairs, it can be expressed as $X = x$, where X is the variable, and x is a possible value. For example, someone's state of being alive can be encoded as follows: let us represent "whether this person is alive" with variable X, appoint 1 to refer to yes and 0 for no, then this person's state of being alive can be expressed as $\mathrm{X} = 1$.

However, this expression is not applicable to situations where the states of affairs are probabilistic, such as whether one will win a lottery after purchasing a ticket, whether vaccination will prevent the infection of a disease, and so on. Strictly speaking, except for the events that have already occurred, which can be viewed as certain, most events are probabilistic in our world. For instance, although our everyday experience suggests that the probability of a person walking through a brick wall seems zero, quantum tunneling theory posits that this probability, while extremely small, is not exactly zero (Ford, 2004, p. 128).

Thus, expression with *probability* allows a more accurate description of more states of affairs. For uncertain cases, suppose a lottery sells 100,000 tickets with ten winning tickets, and a person buys one ticket. The state of affairs regarding whether he will win cannot be expressed using $X = x$, while it can be described using the probability $P(X = 1) = 0.0001$,

where the variable X represents "whether his ticket will win the lottery," with 1 for yes and 0 for no, and 0.0001 is the probability of his winning. For cases with certainty, this expression is also suitable by setting the probability to 1 or 0. For example, it can be expressed as $P(X = 1) = 1$ for someone being alive. Therefore, to some extent, the probabilistic expression $P(X = x) = p$ seems to be a universal way of expressing states of affairs. This is exactly the expression widely used in contemporary theories of causation (Hitchcock, 1993; Woodward, 2016a; Pearl and Mackenzie, 2018, p. 14).

Unfortunately, the expression $P(X = x) = p$ still has limitations, as it does not apply to involving discrete variables with multiple possible values or continuous variables.[1]

- As for discrete variables with multiple possible values, for example, suppose one side of a six-sided die is red, two sides are yellow, three sides are blue, and it is covered by a black box after being rolled. From the observer's perspective, the state of the die (color of the top face) cannot be described with a single $P(X = x) = p$.

- As for continuous variables, take estimating someone's lifespan as an example. Since the probability of any specific lifespan is theoretically close to zero, $P(X = x) = p$ cannot be used to represent an estimate of lifespan.

Then how should these two situations be described by formal expressions? Actually, using '*probability distribution*' can aptly describe them, that is, comprehensively depicting *all possible values* and the corresponding probabilities (for discrete variables) or probability densities (for continuous variables). The basic idea of 'probability distribution' can be found in any probability theory textbook (DeGroot and Schervish, 2012).

For example, in the case of the six-sided die, if X represents the top face color, with 1 for red, 2 for yellow, and 3 for blue, then the state of the top face color can be expressed

[1]Discrete variables are those that take specific, separate values like gender, while continuous variables can take any value within a range, such as temperature, which can vary continuously.

as: "$P(X = 1) = 1/6,\ P(X = 2) = 1/3,\ P(X = 3) = 1/2$," indicating a 1/6 probability of being red, a 1/3 probability of being yellow, and a 1/2 probability of being blue. In the lifespan estimation case, assuming the current age is 50 and the human lifespan limit is 150, with X representing lifespan, then one possible description of the lifespan estimation could be:"$P(50 < X < 150) = 1, P(75 < X < 90) = 0.9$," meaning the lifespan is longer than 50 but less than 150 years, with a 90% probability of being between 75 and 90 years.

Moreover, for expressing 'changes' of states of affairs, the probability distribution expression is more precise and sensitive than the probability expression $P(X = x) = p$. In situations where only a small part of a whole thing changes and other parts remain unchanged, an observer who focuses only on the probability of one of the unchanged parts might mistakenly conclude that the whole thing is unchanged and thus uninfluenced by other things. In contrast, when using the probability distribution expression, any change in the probability of any part will change the probability distribution of the whole, thereby clearly revealing the change of state of the whole. For example, suppose that after a twelve-sided dice is rolled, a cheater can change the top face into 12 only when the result of rolling is 1. If the other observer, in many experiments, only records the probabilities of certain numbers, such as "$P(D = 3),\ P(D = 6),\ P(D = 9)$," and finds that these probabilities do not change, remaining close to 1/12, then this observation might lead him to conclude that the dice's state is normal and unaffected by the cheater. However, if he uses the probability distribution to express the state of the dice, it would reveal the variation, thereby indicating that the dice's outcome is influenced by the cheater.

General speaking, for cases involving discrete variables, the state of affair can be expressed as all possible values and their corresponding probabilities, precisely expressed as:

$$P(X = x_i) = p_i, i = 1, 2, \ldots, n. \quad \sum_{i=1}^{n} P(X = x_i) = 1 \tag{1}$$

Here, p_i is the probability that X takes the value x_i.

For the cases involving continuous variables, the state of affairs can be described as all possible values and the probability density within the corresponding value range, precisely expressed as:

$$P(a \leq X \leq b) = \int_a^b f(x)dx, \quad x \in K, \quad \int_K f(x)\,dx = 1 \tag{2}$$

Here, $P(a \leq X \leq b)$ is the probability of X falling within the interval [a, b], K represents the set of all possible values that X can take, and $f(x)$ is the probability density function of X that assigns probabilities to each value in the set K.

For the cases involving variables whose possible values are partly discrete and partly continuous, the expression of the state of affairs is a combination of the two expressions above, with the sum of probabilities of all values being 1. Cases of certainty (such as events that have already occurred) can be viewed as a special case of discrete distribution, namely $P(X = x) = 1$.

Thus, whether the case is certain or uncertain, the state of affairs can be described using the expression involving probability distribution. That is to say, we can get a theorem as follows.

Theorem 1. *A state of affairs can be expressed by a probability distribution.*

This paper will use $S(X)$ as the sign of the state of affairs referring to the variable X, which comprises two parts: all possible values of a variable and the probability associated with these values (or ranges of values). As a descriptive form of a state of affairs, $S(X)$ can be represented both in an *exact* quantitative manner and in a *fuzzy* manner, adapting to the needs of scientific expression as well as to the needs of everyday cognitive expression. For instance, if a teacher wants to reform teaching strategies to improve student's academic performance, she can roughly estimate the results under different strategies in the sense of probability distribution without needing precise quantification.

The graphical representation of probability distributions also facilitates a simple and intuitive understanding of states of affairs, because a specific probability distribution represents

a specific state of affairs. For example, in the case of teaching strategy reform, students' academic performance under different strategies can be represented as Figure 1, where the y-axis represents students' academic performance, and the $P(Y)$-axis represents the proportion of students at each performance level. The area under the curve above the y-axis represents the 'probability distribution,' i.e., "academic performance of the whole class." Different curves represent different performances under corresponding teaching strategies. Such graphs intuitively show the differences between outcomes. For example, $S_4(Y)$ shows generally higher performance and less disparity among students compared to $S_1(Y)$. This intuitiveness is a clear advantage of representing states of affairs as probability distributions.[2]

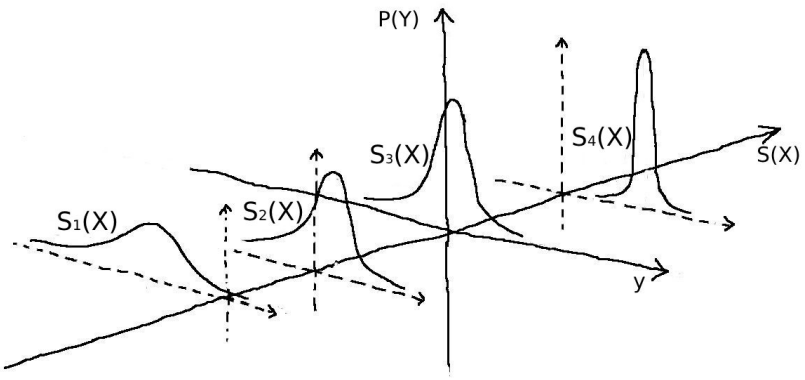


Figure 1: Understanding states of affairs with probability distributions.

Slightly different from the general use of 'probability' in probability theory textbooks, except for special explanations, this article uses this concept in a broader sense, sometimes not clearly distinguishing it from the concepts of possibility or likelihood. There are two main reasons for this.

On the one hand, from the metaphysical perspective, the state of affairs may be objective or subjective, and in both cases, it is possible to describe the state of affairs as a probability distribution. It is crucial to consider the inherent probability of certain events and the limits

[2]This graph is an example of a continuous variable. The expression of discrete variables is simpler in principle, consisting of histograms, as detailed in contemporary probability theory. For brevity, these are omitted here.

of observers' of knowledge. In some cases, the uncertainty of a state of affairs is objective, such as the theoretical outcome of a die roll which has not been done. In other cases, the uncertainty is not objective but reflects the observer's lack of information. For instance, when a die roll has been done and covered by a black box, the outcome is objective because others can check and get the same result, but the observer can only describe it in terms of possibilities. The distinction between subjective and objective states of affairs is also helpful for the in-depth analysis of objective causation and subjective causation, which will be discussed later.

On the other hand, from a cognitive perspective, it is sometimes difficult to distinguish the subjectivity and objectivity of a state of affairs, and it is often not necessary to strictly distinguish between the two in daily life. Because human cognitive activities must be conducted from a subjective perspective, and we cannot know all the information about the external world. In reality, the distinction between these two uses is not always clear. For example, when considering the effectiveness of a vaccine for an individual, even though the outcome may be objectively determined by the vaccine's characteristics and the individual's physiological condition, the outcome is uncertain in the view of a doctor without knowing all that information, and he can only describe it in terms of the probability or likelihood based on the effective rate in the group.

3 Probability distribution determinism: if... then...

Based on our experience, it is incredible to view things in the world as totally disconnected. If we think there are connections between things, then it inevitably involves a certain degree of determinism, or at least local determinism, which means in a given context, specific conditions inevitably lead to certain results (Lewis, 1973; Cartwright, 1979; Gallow, 2016; Woodward, 2016a).

This article does not intend to argue whether the world truly follows determinism com-

pletely, as making a rigorous judgment on this point is likely to exceed the limits of human cognitive ability. What this article is actually concerned with is the logical feasibility of determinism and the rationality of assuming determinism as a way to understand the relationships between things in the world, and they will be demonstrated in this section and the next section.

Typical determinism says that a series of conditions determines the result. Traditional views of determinism often interpreted it in terms of specific results (like $X = x_i$), but this becomes inapplicable when results are uncertain (i.e., $P(X = x_i) = p, 0 < p < 1$). For instance, phenomena like radioactive decay in quantum mechanics can only be described involving probability. According to traditional interpretations of determinism, such systems are non-deterministic. However, a closer examination of the characteristics of these probabilistic outcomes reveals that the probabilistic nature itself possesses a level of stability. For example, probabilistic phenomena in quantum mechanics can be described by a definitive equation, such as the Schrödinger equation. Therefore, phenomena characterized by probabilities in quantum mechanics can be viewed as genuinely deterministic from a certain perspective (Gallow, 2016).

As Theorem 1 argued, states of affairs can be understood as probability distributions, and each state corresponds to a probability distribution with unique identifiable characteristics. Thus, a result of determinism is a specific probability distribution, which can be both certain and probabilistic. That is to say, traditional determinism can be adjusted to accommodate situations of both certainty and uncertainty by interpreting states of affairs with probability distributions. This article defines this updated determinism as follows.

Definition 3.1. ***Probability Distribution Determinism (PDD)****: A series of specific conditions $\mathcal{E}$ can deterministically produce a state of affairs $\mathcal{R}$ which can be represented by a probability distribution, that is, "if $\mathcal{E}$, then $\mathcal{R}$."*

For convenience, this paper denotes the relationship in Definition 3.1 with '$\Longrightarrow$'. For

example, the relationship "if $\mathcal{E}$, then $\mathcal{R}$" can be represented as "$\mathcal{E} \Longrightarrow \mathcal{R}$".[3]

In Definition 3.1, the term used to describe conditions is '*a series of*' instead of 'a set of', this is because the order of the conditions is vital to the outcome. So the conditions include not only the elements but also the order of these elements. This order is metaphysical rather than temporal in the empirical world. The concept of 'a series of conditions' is akin to a 'tuple' in computer programming, where a tuple is a data structure that holds multiple elements in a specific order. This ordered property distinguishes tuples from sets. Sets are unordered and do not allow duplicate elements, whereas tuples maintain the order of elements and can contain duplicate elements.

The core concept of Definition 3.1 is the '*if... then...*' function, and it has almost the same meaning as the fundamental concept in programming and logic, so it does not need further reduction to other concepts for accurate interpretation. In computer programming, the '*if... then...*' function is used to execute a corresponding operation when certain conditions are met: if the condition is true, a specific block of code is executed.

PDD is logically possible and feasible. For instance, in computer programming, the following three programs are conceivable. (For simplicity, the information about the background conditions is omitted.)

- If input 'on', then output the light is on.
- If input 'on', then output the light is on with a 50% probability.
- If input 'on', then output the light is red with a 20% probability, yellow with a 30%

[3]This paper distinguishes between using '$\Longrightarrow$' and '$\rightarrow$'. '$\Longrightarrow$' means 'necessarily leads to', equivalent in logical strength to logical implication, with its subject being specific conditions and states of affairs, e.g., "$S(X), \mathcal{E} \Longrightarrow S(Y)$". '$\rightarrow$' is widely used in other literature to represent causal relationships, e.g., '$X \rightarrow Y$'. In this paper, it is actually a *simplified* expression denoting that the state of X is part of the conditions that determine the state of Y, omitting background conditions $\mathcal{E}$ and the expression of state $S(\ )$.

probability, and blue with a 50% probability.

As mentioned earlier, not only is the third scenario a typical example of PDD, but the other two cases can also be represented using probability distributions.

The new definition of determinism emphasizes the sufficiency of the relationship. Therefore, the conditions must comprehensively include information that is traditionally treated as background conditions. Even if it cannot be clearly described, it needs to be vaguely referred to, for instance, with statements like "with some other conditions." This is because that it is logically possible to imagine a scenario where one of the background conditions may change, thereby affecting the result, and in explaining such a situation, that condition cannot be ignored. This can be understood through the following scenario: for a programmer of a computer program, any parameter which is part of the conditions determining the outcome of the program can be manipulated to produce a different result, thus this parameter can not be ignored when we explain the result.

For events in our world, all its conditions include the fundamental physical properties of the universe. Because any fundamental physical property of the real world could appear in a different way from our world, leading to different states of the things in our world. For example, in our universe, we take for granted the homogeneity of space which means that every point in space is equivalent to every other, and thus the laws of physics are the same in one place as another. We also believe in direction invariance which refers to the property where a physical law or phenomenon remains unchanged under different spatial orientations. However, these principles are not necessary but only because it happens to be a feature of the universe we live in, and our experience with the world around us conditions us to believe it. Theoretically, one could imagine hypothetical universes with heterogeneous space and preferred directions—universes in which the laws of nature depended on location and direction, where physicists would need to figure out not only the laws of physics but the rules about how those laws change when location and direction change. (Ford, 2004, pp. 157-158).

Therefore, the relationships between things in PDD are mostly context-dependent rather

than absolute. From the perspective of PDD, the scientific knowledge we use to describe the external world is relative to conditions such as a series of fundamental properties of our universe. Suppose there is another universe with fundamentally different properties from our world, then our scientific knowledge may not hold true there. For example, Newton's law of universal gravitation, which is expressed as $F = G\frac{m_1 m_2}{r^2}$ and states that the force of gravity between two masses is directly proportional to the product of their masses and inversely proportional to the *square* of the distance between their centers, should be strictly stated as, "In a universe like ours, $F = G\frac{m_1 m_2}{r^2}$." Because it is possible that there is another higher-dimensional universe where gravity is inversely proportional not to the *square* of the distance (r^2) but to the *cube* of the distance (r^3), that is to say the law of universal gravitation in there might be expressed as $F = G\frac{m_1 m_2}{r^3}$.

On that ground, the stability of background conditions determines the stability of the relationship between things. Only when background conditions are stable, do relationships between things show stable regularity (like physical and chemical knowledge), and when background conditions are unstable, they do not constitute a regularity (like some epidemiological or sociological knowledge). Therefore, the concept of causation which involves determinism can not be simply defined with 'regularity' as proposed by Hume (2007, p. 69). Moreover, for the problem of external validity of causal knowledge whether the conclusion on one occasion is transferable or valid in another occasion, it essentially depends on whether the background conditions are similar in these contexts.

Definition 3.1 only requires sufficiency, which means the condition may not be necessary for the result. This can be understood as a kind of *asymmetric* mapping relationship, i.e., a series of specific conditions $\mathcal{E}_1$ corresponds to a unique result $S_1(Y)$; however, a result may correspond to multiple different series of conditions, i.e., another series of conditions $\mathcal{F}_1$ may

also produce the same result $S_1(Y)$. That is,

$$\begin{cases} \mathcal{E}_1 \Longrightarrow S_1(Y) \\ \mathcal{F}_1 \Longrightarrow S_1(Y) \end{cases} \tag{3}$$

This asymmetric relationship can be understood with this example: it is possible that two software companies receive the same task from a client and then write two programs independently. The codes of these two programs differ in many details, but they produce the same result. Similarly, we can imagine another universe with completely different basic physical conditions from ours, yet having similar macroscopic state results. Furthermore, this occasion can happen not only at the overall level but also at the local level. For example, under certain conditions $\mathcal{E}_0$, taking either drug A or drug B can lead to the cure of someone's disease $S_0(H)$. Or, for computer programming, there may be two series of subprograms, and the use of either can lead to the same result.

The focus on sufficiency rather than necessity makes the PDD system compatible with things like *free will*. '$\mathcal{E} \Longrightarrow \mathcal{R}$' emphasizes that conditions can lead to results, rather than the dependence of results on conditions, that is, it is not equivalent to 'if $\mathcal{R}$ exists, then there must be some conditions $\mathcal{E}$', which is the thinking perspective of some traditional determinism. Thus, under Definition 3.1, there could exist things whose states are not entirely determined by anything else, while they can be parts of the conditions which can produce certain outcomes. And this can be regarded as the definition of free will under the PPD framework. It can be expressed as follows.

$$\begin{cases} \quad \ldots \not\Longrightarrow \mathcal{FW} \\ \mathcal{FW}, \mathcal{E} \Longrightarrow \ldots \end{cases} \tag{4}$$

Where $\mathcal{FW}$ refers to free will, $\not\Longrightarrow$ means 'cannot determine', and $\mathcal{E}$ refers to some other conditions.

The key point of 'free' is not that there is no limit of the range of options, but that even if there is a limited range of options, the result of it is not completely determined by any other conditions. For instance, in computer programming, the options of an input signal may be finite, but the result of the input is not determined by the internal information of the program system.

If the physical world in which humans live is regarded as a completely deterministic system, we can say that free will is out of this system, the state of it is not completely determined by the information of the physical world, and it can input information into the physical world to influence the things in the system. This statement is obviously compatible with the religious claim that the soul with free will is an existence outside this physical world. The discussion here is only to show that PDD does not exclude this claim, rather than support it.

Since Section 2 distinguishes the objective level and the subjective level for the state of affairs (probability distribution), we can distinguish between objective effects and subjective effects. Suppose that at the objective level, the state (probability distribution) of an event remains unchanged regardless of whether an action is taken or not. However, from an individual's subjective perspective, due to not having access to all the information of conditions, he may perceive a risk for the result if he does not act, and thus he decides to take action to reduce the risk. In such a scenario, it can be said that his action has no effect objectively, but it does have an effect from the subjective perspective of the person.

Take vaccination as an example. Objectively speaking, based on an individual's physical characteristics and the characteristics of the virus, it's possible that even without vaccination, the person would not contract the disease; hence, vaccination does not influence their likelihood of falling ill. However, subjectively, due to a lack of understanding about relevant information, the individual is unable to know whether he will get ill if not getting vaccinated. He might only believe that the probability of contracting the disease without vaccination is much higher than with it. Therefore, from a subjective standpoint, getting vaccinated is

effective, as it reduces the perceived risk of illness.

4 Similar to reality: individualization and nestability

When understanding relationships, we usually focus on individualized things. However, the extent to which something is considered an 'individualized' thing is not purely objective but depends on our cognitive needs, which can vary depending on the situation. What may be considered as an individualized object in one context may be perceived as a mix of lots of different things in another context. In certain contexts, we may focus on specific attributes of an object and make it stand out as an individual, while in other situations, we might group objects based on shared characteristics, downplaying their individual distinctions. This cognitive flexibility highlights the dynamic nature of individualization, revealing how our perception of individuality is adaptable and context-dependent.

The flexibility of individualization is reflected in the activities of naming and referring. We can use various names for different objects, or we can view these objects collectively and refer to them by a single name. This section will demonstrate that, relying solely on PDD and the strategy of naming, we can derive systems with complex relationship structures akin to those in the real world.

The fundamental form of relationships in PDD is '$\mathcal{E} \Longrightarrow \mathcal{R}$', so the relationships named in an integrated manner within this system must be based on this and have additional elements. Thus, there are initially two basic ways of integration: one is the synthesis of multiple parallel determinative relationships into 'one' determinative relationship, and the other is the integration of multiple successive determinative relationships into a simplified determinative relationship. Additionally, let us call $\mathcal{E}$ in the determinative relationship an effective condition, and other information that does not affect this process is deemed ineffective, then there is a possible way of integration, namely, the combination of ineffective and effective information to form an 'individualized' condition. These three methods can be

combined to form more complex ways of integration. Below, the systems under these three basic integrated naming strategies will be analyzed in detail.

Firstly, let us look at the synthesis of multiple parallel determinative relationships. Suppose there are some relationships as follows,

$$\begin{cases} \mathcal{A}_1, \mathcal{E} \Longrightarrow \mathcal{B}_1 \\ \mathcal{A}_2, \mathcal{E} \Longrightarrow \mathcal{B}_2 \\ \quad \vdots \\ \mathcal{A}_n, \mathcal{E} \Longrightarrow \mathcal{B}_n \end{cases} \tag{5}$$

If we view $\mathcal{A}_i, \mathcal{A}_j$ as a whole, collectively referred to as $\mathcal{C}_m$, and $\mathcal{B}_i, \mathcal{B}_j$ as a whole, collectively referred to as $\mathcal{D}_m$,

$$\begin{cases} \mathcal{C}_m \equiv \{\mathcal{A}_i, \mathcal{A}_j\} \\ \mathcal{D}_m \equiv \{\mathcal{B}_i, \mathcal{B}_j\} \end{cases} \tag{6}$$

then

$$\mathcal{C}_m, \mathcal{E} \Longrightarrow \mathcal{D}_m \tag{7}$$

Again, we can view $\mathcal{C}_i, \mathcal{C}_j$ as a whole, collectively referred to as $\mathcal{G}_m$, and view $\mathcal{D}_i, \mathcal{D}_j$ as a whole, collectively referred to as $\mathcal{H}_m$,

$$\begin{cases} \mathcal{G}_m \equiv \{\mathcal{C}_i, \mathcal{C}_j\} \\ \mathcal{H}_m \equiv \{\mathcal{D}_i, \mathcal{D}_j\} \end{cases} \tag{8}$$

then

$$\mathcal{G}_m, \mathcal{E} \Longrightarrow \mathcal{H}_m \tag{9}$$

And so on, many nested layers can be formed based on this naming strategy.

Thus, (5) can be seen as a detailed explanation of (7), and (7) can be seen as a detailed explanation of (9).

In our world, there are many similar cases, for example, a compound medicine ($\mathcal{C}_m$) containing multiple active ingredients ($\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$), each of which has an independent effect that can influence different aspects of the body ($\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n$). The integration of these effects is the alleviation of a person's illness ($\mathcal{D}_m$). In this context, the effects of each component ($\mathcal{A}_i, \mathcal{E} \Longrightarrow \mathcal{B}_i$) serve as a micro-level explanation of the medicine's ability to cure diseases ($\mathcal{C}_m, \mathcal{E} \Longrightarrow \mathcal{D}_m$).

Secondly, let's examine the integration of multiple successive determinative relationships. Suppose there are a series of relationships as

$$\begin{cases} \mathcal{A}_1, \mathcal{E}_1 \Longrightarrow \mathcal{A}_2 \\ \mathcal{A}_2, \mathcal{E}_2 \Longrightarrow \mathcal{A}_3 \\ \qquad \vdots \\ \mathcal{A}_n, \mathcal{E}_n \Longrightarrow \mathcal{A}_n \end{cases} \tag{10}$$

where each $\mathcal{E}_i$ does not affect the others relationships.

We can collectively refer to $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n$ as $\mathcal{F}$,

$$\mathcal{F} \equiv \{\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_n\} \tag{11}$$

then (10) can be expressed as:

$$\text{conditional on } \mathcal{F}: \quad \mathcal{A}_1 \Longrightarrow \mathcal{A}_2 \Longrightarrow \mathcal{A}_3 \Longrightarrow \ldots \Longrightarrow \mathcal{A}_n \tag{12}$$

Due to the inevitability of determinative relationships, even if some parts are omitted, the

relationship still holds,

$$\text{conditional on } \mathcal{F}: \quad \mathcal{A}_1 \Longrightarrow \mathcal{A}_3 \Longrightarrow \mathcal{A}_5 \Longrightarrow \ldots \Longrightarrow \mathcal{A}_n \tag{13}$$

$$\text{conditional on } \mathcal{F}: \quad \mathcal{A}_1 \Longrightarrow \mathcal{A}_5 \Longrightarrow \mathcal{A}_9 \Longrightarrow \ldots \Longrightarrow \mathcal{A}_n \tag{14}$$

$$\vdots$$

$$\text{conditional on } \mathcal{F}: \quad \mathcal{A}_1 \Longrightarrow \mathcal{A}_n \tag{15}$$

Regarding such a series of relationships, (14) can be seen as a detailed explanation of (15), while (13) can be considered a detailed explanation of (14), and so on. In other words, they can be viewed as relationships at different levels of granularity, constituting multiple nested hierarchies.

There are also many similar nested hierarchical relationships in the real world. For example, the macroscopic effects of drugs on diseases can be detailed as a series of microscopic processes. First, the drug is absorbed into the bloodstream. Then, it is transported throughout the body by the blood, particularly to its targeted tissues or organs. Subsequently, the drug interacts with specific cellular receptors or enzymes at the target site, initiating biochemical reactions that lead to its therapeutic effect.

Thirdly, let us consider the combination of ineffective and effective information. Suppose there is a relationship as

$$\mathcal{A}, \mathcal{E} \Longrightarrow \mathcal{C} \tag{16}$$

and $\mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3, \ldots, \mathcal{U}_n, \ldots$ are things that does not affect the relationship between $\mathcal{A}$ and $\mathcal{C}$, then,

$$\mathcal{A}, \mathcal{E}, \mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_3, \ldots, \mathcal{U}_n, \ldots \Longrightarrow \mathcal{C} \tag{17}$$

If we view $\mathcal{A}$ and $\mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_n$ as a whole, denoted as $\mathcal{B}_n$,

$$\mathcal{B}_n \equiv \{\mathcal{A}, \mathcal{U}_1, \mathcal{U}_2, \ldots, \mathcal{U}_n\} \tag{18}$$

then

$$\begin{cases} \mathcal{B}_1, \mathcal{E} \Longrightarrow \mathcal{C} \\ \mathcal{B}_2, \mathcal{E} \Longrightarrow \mathcal{C} \\ \quad \vdots \\ \mathcal{B}_n, \mathcal{E} \Longrightarrow \mathcal{C} \\ \quad \vdots \end{cases} \tag{19}$$

and for any $i < j$, $\mathcal{B}_i$ is a part of $\mathcal{B}_j$.

In this system, $\mathcal{B}_i$ and $\mathcal{B}_j$ are in a part-whole relationship, and both can lead to the same result. Thus, $\mathcal{B}_i, \mathcal{E} \Longrightarrow \mathcal{C}$ can be seen as the refinement of $\mathcal{B}_j, \mathcal{E} \Longrightarrow \mathcal{C}$.

Relationships with this structure are also common in the real world. One typical example is the relationship between citrus fruits, vitamin C, and scurvy. Historically, it was empirically discovered that consuming citrus fruits could prevent scurvy. As medical science advanced, it became clear that the critical ingredient within citrus fruits responsible for this benefit was vitamin C rather than anything else like acidity. Therefore, the intake of vitamin C leading to the prevention of scurvy is a refined understanding of the initial broader observation on the relationship between citrus fruits and scurvy. It can be shown graphically as follows:

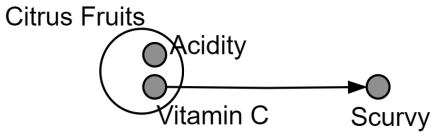
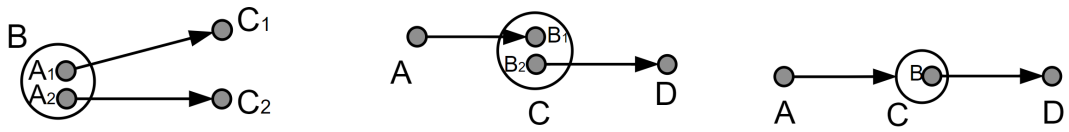


Figure 2: Relationship between citrus fruits, vitamin C, and scurvy.

Pearl failed to properly handle such nested relationships, he considered vitamin C as a mediator in a causal chain and denoted the relationship as "Citrus Fruits → vitamin C → Scurvy" (Pearl and Mackenzie, 2018, p. 300). This treatment is inappropriate because there is not a determinative relationship or causal relationship between Citrus Fruits and vitamin C, but a *whole-part* relationship. Thus, a shortcoming of Pearl's theory is its neglect of such whole-part relationships when analyzing relationships between things using *causal diagrams*,

leading to the misrepresentation of non-causal relationships as causal ones.

Relationships between things are not limited to causal ones but also include non-causal ones, such as the relationship between a whole and its parts. Therefore, to accurately represent the connections between things in causal diagrams, it is essential to consider the presence of non-causal relationships. Additionally, in different situations, we often need to consider the relationships between things at different levels. When multiple local causal networks at different levels are incorporated into a large causal network, it may appear that the two individualized things in local networks are different parts of an individualized thing in another level. These relationships may be like the below:



(a) (b) (c)

Figure 3: Causal networks involving whole-part relationships.

The treatment of whole-part relationships helps us to better grasp the calculation of causal effects. Some other theories of causation are ambiguous on issues such as the additivity of causal effects (Pearl and Mackenzie, 2018, p. 326) and the transitivity of causal relationships (McDonnell, 2018). Now we can at least deepen our understanding of these issues: within causal networks, for segments involving whole-part relations, the transitivity of causal relationships may be interrupted, and the calculation of the overall causal effect needs to consider both the effect size of the causal part and the relation between the part and the whole.

The universality of the above three nested systems in the real world is more fully reflected in the mechanistic theory of causation. The analyses of the characteristics of scientific knowledge by Glennan (1996, 2002) and Machamer et al. (2000) suggest that, except for possible fundamental processes, almost all causal relationships are typically underpinned by complex systems, which produce those behaviors through the interaction of numerous

parts. Those systems are referred to as mechanisms. Mechanisms exhibit nested hierarchical structures with multiple layers. In these hierarchies, the connections between various levels are characterized by part-whole relationships, and the processes at the lower levels form the basis for the emergence of phenomena at the higher levels (Machamer et al., 2000). According to this, it can be said that (5), (7) are the mechanisms of (9), while (12), (13), and (14) are the mechanisms of (15).

From the above analyses, we can get this:

Theorem 2. *Based solely on Probability Distribution Determinism and naming strategies, it is possible to deduce systems whose nested hierarchical structures are highly similar to those found in the world we inhabit.*

This supports the legitimacy of PDD as a basis for studying the relationships between real-world events.

Furthermore, these analyses can provide insights for scrutinizing some other theories of causation, such as mechanistic theory, structural equations, and ontological interpretations of causal relationships.

Mechanistic theory undoubtedly helps us understand the relationships between things more deeply. One significant insight is that the existence of a micro-level mechanism can provide a strong rational basis for the hypothesis that there is a causal relationship between things at the macro level. Causal claims with mechanistic explanations are usually more credible than those lacking such explanations.

However, there are issues with the mechanistic theory's definition of the concept of causation. As for the priority of the concepts of mechanism and causation, it proposes that mechanisms precede causation, which means that mechanisms underlie most causal connections and do the causal work (Glennan, 2002, 2010). This is refuted by the new theory. Firstly, this assertion itself has an internal logical flaw because it cannot answer whether the description of the local details of a mechanism can avoid the concept of causation. Logically speaking, a mechanism cannot be infinitely reduced to a more microscopic mechanism, or,

one can imagine a relationship without microstructure. For example, in a computer program, the relationship between actions and effects might be merely based on the if-then function, lacking a subdividable structure, and therefore no mechanism exists. A character 'hit' by a 'bullet' in the 'head' might be programmed to 'die' instantly, without certain parts of the 'brain' and 'nerves' being damaged, which leads to the disruption of 'life systems.' In this situation, it is obvious that the basic irreducible relationship cannot be described by the concept of mechanism, but it can be described by the concept of determination in PPD, which is related to the concept of causation. Secondly, according to the analysis in this paper, a system similar to the mechanism in the real world can be built based on PDD, and the relationship in PDD involves the concept of causation (see Section 7 for details). Therefore, it is more reasonable to regard the concept of causation as the basis of mechanism than the contrary claim held by the mechanistic theory.

Regarding the debates about the *ontological explanation* of causation, the analysis of this paper can also provide insights. In mechanistic explanations, the basic elements of causal relationships are objective entities and activities (Machamer et al., 2000). *Transfer theories* view the basis of causal relationships as involving the transfer of energy, momentum, electrical charge, or information, based on objective real connections, or consider causal relationships to have certain spatiotemporal properties, such as spatial adjacency and temporal sequence (Mumford and Anjum, 2013, p. 66). However, causal relations are not necessarily limited to the physical level; they can also be abstract, such as how gender characteristics influence hiring in a company with gender discrimination. The analysis of this paper shows that relationships of one thing influencing the other may not depend on physical characteristics. A more intuitive example comes from virtual reality in computer programs. One of the bases of computer programming is the if-then function. Programmers use these functions to write codes, ultimately constructing programs that closely resembles the real world in terms of phenomenological representation. In computer programs, there are also relationships of influence between events or phenomena, and this concept of 'influence' can also be regarded

as 'causation' (Wheeler, 2022). In the virtual world of computer programs, things with physical characteristics do not exist, and the fundamental basis for the 'relationships of influence' between phenomena is the if-then functions in the code. Therefore, causation can be based solely on if-then function without needing other ontological assumptions.

As for whether *temporality* is a necessary attribute of the concept of causation, this paper also provides an interpretation different from traditional theories. A challenge to theories that define causation based on temporal sequence is the phenomenon of quantum entanglement, where a change in the state of one entangled quantum instantly affects the state of another, without any temporal sequence (Ford, 2004, p. 231). Such scenarios are difficult to explain using causality theories reliant on temporality. Within the framework of this paper, on the one hand, relationships based on PDD does not necessarily require temporality; on the other hand, the asymmetry of relationships based on PDD is compatible with the asymmetry of temporal sequence. Thus, PDD can explain both relationships involving temporal sequence and those that do not. In the relationship between conditions and outcomes in a computer program, there exists a purely logical and non-temporal relationship. Specifically, a series of parameters, functions, and input signals constituting all conditions determine the outcome of the program in an absolute logical sense. This determinative relationship does not depend on temporality but merely on the program's logical sequence. Even if not executed and merely written on paper, the same results can be necessarily deduced from the defined program parameters and operating rules as those obtained from actual execution.

The theory of causation asserting that the cause must precede the effect faces certain difficulties in explaining the following hypothetical scenario: Suppose a primitive man from ancient times time-travels to the modern era. He is unaware that weather can be predicted. Based on his repeated observations, he notices that whenever there are many people carrying umbrellas, it subsequently rains. If he adheres to the theory of causation mentioned above, he might deduce that carrying umbrellas causes rain (he may even think that the collective act is a form of rainmaking witchcraft practiced by modern people). Obviously this is wrong.

More generally, it is conceivable that someone might have the ability to foresee the future. For a person with foresight, his or her present actions could be affected by future events. In this scenario, the cause is temporally subsequent, whereas the effect is temporally antecedent. For instance, suppose someone foresees that a bridge will collapse, leading he to choose a different route instead of crossing the bridge. Here, the bridge's collapse occurs later in time, while the person's change of route occurs earlier. In terms of causal explanation, it is because the bridge will collapse that he changes his route, that is, the cause is subsequent, and the effect is antecedent. Since PDD does not require temporality, the new theory in this article will not be challenged when facing such situations.

The above analysis may have greater implications for reflecting on the expressive ability and application scope of the *structural equation* method in representing causal relationships. Currently, in the theoretical domain of causality and in the social and behavioral sciences, structural equations are quite popular for characterizing the relationships between things. For instance, an equation like $Y = f(X, U_Y)$ is used to represent the knowledge that "X can influence Y," where U_Y stands for unobserved variables that affect Y. The equation is 'structural' in that variable Y does not enter into the equation about causes of X (Pearl and Mackenzie, 2018, pp. 277-278). However, according to the previous analysis, this method of expression faces multiple challenges. Firstly, since any background condition could theoretically change and thereby affect the outcome, it seems that all determinative conditions should be included in the equation, but that is infeasible. Secondly, due to possible naming approaches and various cognitive levels, there are infinite conditions that can affect the outcome. Therefore, structural equations cannot be perfect expressions of causal relationships.

For practical purposes, maybe it is unnecessary to include every possible influencing condition in the equation, and the causal knowledge expressed by the equation could be considered valid only in specific situations, not absolutely. However, this approach raises a new issue: how do we decide which variables to include? By what criteria should we determine the types of objects that can be incorporated into the equation (Woodward, 2016b)? Perhaps

an intuitive judgment is that the variables need to be at the same level in the aforementioned nested hierarchies, but this knowledge itself might transcend empirical cognitive abilities and be indeterminable. In other words, the rationality of selectively including variables, as well as the rationality of the criteria for variable selection, both require further in-depth analysis and demonstration.

5 From general relations to specific events

Determinism describes relationships in a general sense, however, even if something is one of the conditions that can determine a result, it is not always considered one of the factors for the explanation of a specific event. When explaining everyday events, people do not consider the basic physical properties of our universe, even though they are fundamental supportive conditions for those events. That is, there are two different questions:

- Q1: In the general sense, what conditions can produce a specific result?
- Q2: In a specific situation, what is responsible for the change of the state of a thing?

As practitioners in a dynamic world, when we interpret events in real life, we mostly do it in a manner similar to Q2. Clearly, not all supportive conditions for a specific event are considered factors in an explanation. For example, the presence of air is one of the conditions for fire, but it is usually not regarded as the responsible factor of a fire. However, is the presence of air always not considered a cause of fire (responsible for fire accidents)? Not necessarily. Consider the following two scenarios involving air and fire:

- F1: Suppose a forest fire occurs. Should the presence of air be considered one of the factors responsible for the fire?
- F2: Suppose there is some oil in a high-temperature vacuum tank. Since there is no air in the tank, no fire occurs, and this state continues for a long time. Then, at a

certain moment, someone injects air into the vacuum tank, and a fire starts. Should the presence of air be considered the factor responsible for the fire in this context?

Intuitively, in F1, we usually do not consider the presence of air as one of the factors responsible for the forest fire, but in F2, the introduction of air is considered the factor responsible for the fire, while the presence of oil and the high-temperature condition may not be seen as responsible for the fire. This shows that for specific events in Q2, whether a condition is considered as its cause (responsible for the event's occurrence) does not entirely depend on the general attributes of the condition or its general relationship with the result. Therefore, we can get:

Theorem 3. *Regarding the relationships of influence between things, there are two different contexts: one focuses on the potential relationships of influence in the general sense, and the other focuses on the responsibility for changes of states of affairs in specific situations.*

However, what makes something considered responsible for a specific event? In F1, air is not considered as the cause of the fire because air is always present, and its existence does not differ between the fire and non-fire scenarios. In F2, the presence of oil and high temperature remains constant before and after the fire, while the only difference is the presence of air, so the appearance of air is seen as the responsible party for the event. That is, whether the presence of air is considered the cause of the fire depends not only on the general potential determining relationship between air and fire but also on whether the condition of air's presence changes before and after the fire.

Whether explicitly stated or implied, for events that have occurred, the analysis of the factors responsible for the result actually involves a comparison with the scenario where the phenomenon did not occur. Compared to Q1, which is concerned with all supportive conditions in an absolute sense, Q2 is more interested in what specific variation of condition ($\Delta S(X)$) is responsible for the variation of the state of the other thing ($\Delta S(Y)$).

Theorem 4. *For a specific event, the factors responsible for it must be something whose state has a variation in this situation.*

Based on the analysis above, we can discern the predicaments faced by the counterfactual theory of causation (Lewis, 1973). The basic idea of counterfactual theory is that causal claims are interpreted in terms of counterfactual dependence, formulated as "If A had not occurred, C would not have occurred; A occurred and C occurred." A major challenge for this theory is that one thing may have a counterfactual dependency relationship with another thing but there is no causal relationship between them. For example, suppose the flowers in my house died. Obviously, if a stranger who had never known about the existence of my flowers had watered them, they would not have died. Theoretically, the stranger's action of watering the flowers and the death of the flowers constitute a counterfactual dependency relationship. However, usually, we would not attribute the death of the flowers to the stranger (Menzies and Beebee, 2020). Thus, counterfactual theory apply at most to relationships of influence in the general sense (i.e., the scenario in Q1), and cannot apply to the analysis of factors responsible for a specific event (i.e., the scenario in Q2). According to Theorem 4, the stranger's behavior has no variation when the flowers alive and dead, so it should not be considered the cause of the flowers' death.

6 Epistemological perspective

The previous text analyzes the characteristics of the PDD system from a metaphysical perspective and considers it as a viewpoint for understanding causal relationships. However, merely based on this knowledge, we are unable to determine whether a certain relationship belongs to PDD. Therefore, we also need to analyze this type of relationship from an epistemological angle, so that we know how to identify this relationship without background knowledge.

Two types of relationships under different situations are distinguished in Theorem 3,

and the relationship in the special situation is built upon the relationship in the general sense with additional conditions specified in Theorem 4. Thus, we only need to analyze the general one to understand both of them. The strategy of this section is to assume this kind of relationship and then gradually transform the scenario with metaphysical knowledge into a scenario without prior knowledge. Since in reality we are concerned with the relationship between two specific things, our hypothesis is as follows:

Assumption 6.1. *The state of X can influence the state of Y under certain condition $\mathcal{E}$, that is,*

$$S(X), \mathcal{E} \Longrightarrow S(Y) \tag{20}$$

This relationship can be equivalently expressed as: X has at least two different states, which under the unchanged condition $\mathcal{E}$, can deterministically lead to two different states of Y, that is,

$$\begin{cases} S_1(X), \mathcal{E} \Longrightarrow S_1(Y) \\ S_2(X), \mathcal{E} \Longrightarrow S_2(Y) \end{cases} \tag{21}$$

where $S_1(X) \not\equiv S_2(X)$, $S_1(Y) \not\equiv S_2(Y)$.[4]

The above two expressions are logically equivalent, they imply each other:[5] (21) $\Leftrightarrow$ (20).

Suppose there are various pieces of information $\mathcal{N}$ that have no impact on the above relationships. According to Definition 3.1, conditional information allows redundancy, because even when unrelated information is added to the condition, it can still sufficiently determine the outcome. So when $\mathcal{N}$ is added to the conditions, the determinative relationship remains valid. In the context of computer programs, this information corresponds to redundant code that, due to programmer error, is written into the program but is never executed and has

[4]'$\not\equiv$' means not the same. The reason for not using '=' is that the subject of the formula is 'state'/'probability distribution', which may not necessarily be a single value.

[5]In this paper, $\Leftrightarrow$ is used in a logical sense, meaning mutual logical implication, each can be deduced from the other. It is not used in the sense of determinism.

no impact on the final result of the program. Therefore, we have:

$$\begin{cases} S_1(X), \mathcal{E}, \mathcal{N} \Longrightarrow S_1(Y) \\ S_2(X), \mathcal{E}, \mathcal{N} \Longrightarrow S_2(Y) \end{cases} \tag{22}$$

Obviously, (22) $\Leftrightarrow$ (21).

Regarding what should be included in $\mathcal{N}$, caution is needed, especially concerning intermediate variables. Suppose under certain conditions, $S(X)$ can lead to $S(M)$, and $S(M)$ can lead to $S(Y)$,

$$\textit{conditional on } \mathcal{E} : S(X) \Longrightarrow S(M) \Longrightarrow S(Y)$$

Suppose there is already a certain amount of M in the environment of $S(X)$, we can denote this part as M_0. Off cause, M_0 can independently lead to a part of $S(Y)$. However, the presence of M_0 does not affect the relationship between $S(X)$ and $S(Y)$, so this part of M_0 belongs to $\mathcal{N}$. Since $S(X)$ itself can produce $S(M)$, we cannot say at the variable level (or type level) that M belongs to $\mathcal{N}$.

Suppose there is something can potentially influence $S(X)$, then when $S(X)$ is already determined as $S_1(X)$ or $S_2(X)$, the presence of these things will not affect the relationship between $S(X)$ and $S(Y)$. If we call all these things that can potentially influence $S(X)$ as $\mathcal{I}$, then

$$\begin{cases} S_1(X), \mathcal{E}, \mathcal{N}, \mathcal{I} \Longrightarrow S_1(Y) \\ S_2(X), \mathcal{E}, \mathcal{N}, \mathcal{I} \Longrightarrow S_2(Y) \end{cases} \tag{23}$$

This brings us to a key concept in this theory, the meaning of "when $S(X)$ is already determined", which is a metaphysical concept rather than an empirical one. Although this scenario can be understood through the concept of intervention, like Pearl's explanation that intervention can make $S(X)$ immune to other influences ($\mathcal{I}$) and only subject to the intervention (Pearl and Mackenzie, 2018, p. 41). However, there may be some differences between them. Traditional concepts of intervention are to some extent empirical concepts; the

realization of intervention effects itself depends on certain conditions involving the concept of causation. This is one of the main reasons for the non-reductive nature of interventionism (Woodward, 2016a). In this article, this state is in the metaphysical sense. Drawing an analogy from computer programming, it does not intervene in the target object through certain objects within the program. Instead, from the programmer's perspective, it involves directly modifying the program parameters, resulting in the state of the target object being directly specified. Alternatively, due to hardware reasons external to the program, the parameters related to the state of X are forcibly set, ensuring that no other input information can alter X's state. In the PDD framework, the basis for the validity of the relationship between $(S(X), \mathcal{E}, \mathcal{N})$ and $S(Y)$ in (22) is necessary, and when $(S(X), \mathcal{E}, \mathcal{N})$ is determined, regardless of what $\mathcal{I}$ is, it necessarily determines $S(Y)$.

Equation (23) can be represented as another form:[6]

$$S(Y|S_1(X), \mathcal{E}, \mathcal{N}, \mathcal{I}) \not\equiv S(Y|S_2(X), \mathcal{E}, \mathcal{N}, \mathcal{I}) \tag{24}$$

We can collectively refer to $\mathcal{E}$, $\mathcal{N}$, $\mathcal{I}$ as Ψ, and (24) can be written as:

$$S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi) \tag{25}$$

At this point, we reach a crucial step. $\mathcal{E}, \mathcal{N}, \mathcal{I}$ are not distinguished in Ψ, and they include all related and unrelated things, constituting a context with no need for prior knowledge. This means that aside from the things to be examined, we only know that in the two scenarios being compared, all other information is the same. However, we do not know among these pieces of information which might have an impact on $S(Y)$ and which do not. This is precisely the judgment environment we wish.

[6]'|' represents conditional on. This expression is inspired by the conditional probability notation $P(Y|X)$ in probability theory, which represents the probability of Y occurring given the existence of X.

The content of Ψ can be defined using the common attributes of $\mathcal{E}$, $\mathcal{N}$, and $\mathcal{I}$. According to (23), a common feature of the three is that they are *all* states of affairs that can coexist with $S_1(X)$ and $S_2(X)$. In fact, this is the only information we can obtain about them. While it seems that $\mathcal{E}$, $\mathcal{N}$, and $\mathcal{I}$ can also coexist with $S_1(Y)$ and $S_2(Y)$, they are only parts of not all states of affairs that can coexist with them. That is, a state of affair that can coexist with $S_1(Y)$ and $S_2(Y)$ may not coexist with $S_1(X)$ and $S_2(X)$. Thus, (25) implies a non-symmetrical relationship between X and Y. Since the content of $\mathcal{N}$ is defined at the individual level rather than the type level of the variable, the definition of Ψ should also be at the individual rather than the type level of the variable. That is, for any ψ_i, treated as an individual or property of the individual rather than a variable, if it can coexist with both $S_1(X)$ and $S_2(X)$, then it belongs to Ψ.

Since (25) $\Leftrightarrow$ (24) $\Leftrightarrow$ (23) $\Leftrightarrow$ (22) $\Leftrightarrow$ (21) $\Leftrightarrow$ (20) $\Leftrightarrow$ Assumption 6.1, we can conclude that (25) $\Leftrightarrow$ Assumption 6.1, meaning (25) and Assumption 6.1 can logically imply each other. Therefore, the latter can serve as a definition for the former. That is,

Theorem 5. *$S(X)$ can influence $S(Y)$ under certain conditions if and only if: for two different states $S_1(X)$ and $S_2(X)$ of X,*

$$S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$$

where Ψ represents a series of conditions, and if ψ_i, which is not a variable treated as a type but refers to an individual or property of an individual, can possibly coexist with both $S_1(X)$ and $S_2(X)$, then $\psi_i \in \Psi$.

This expression, when understood in a non-technical way, means that if the state of Y varies when the state of X differs, while all other conditions are the same, then we can say that the state of X can influence the state of Y under those conditions. This expression has a high degree of intuitive fit with everyday intuition, making it not only precise for scientific or philosophical understanding but also applicable to everyday reasoning in a simple and

intuitive way. For example, a gust of wind blows, and a girl's hat moves. In her observation, with wind $S_1(X)$ and without wind $S_2(X)$, almost all other events are identical Ψ. Without wind, her hat doesn't move, which is $S(Y|S_1(X), \Psi)$; but with wind, it does, which is $S(Y|S_2(X), \Psi)$. Therefore, according to Theorem 5, she can infer that the presence of wind caused the hat to move. This shows that Theorem 5 can be used to guide and explain everyday causal reasoning processes.

As for why '*coexistence*' can serve as a foundational concept for causation, firstly, this attribute of 'coexistence' is directly contained in the above formal expression and is even the only attribute we can derive from the above analysis framework. Secondly, coexistence may be more fundamental than causation, as causation is abstract and even polysemous, while coexistence is relatively concrete. Although in the above, 'coexistence' is used in a metaphysical sense, it can also be directly observed and perceived at an epistemological level.

In some popular theories of causation, there are expressions that have a formal similarity to Theorem 5, such as using $P(Y|X) \neq P(Y|\neg X)$ (Cartwright, 1979; Hitchcock, 1993; Rubin, 2005; Woodward, 2016a) or $P(Y|do(X)) \neq P(Y|do(\neg X))$ (Pearl and Mackenzie, 2018, p. 151) to determine causal relationships. That is, if the probability of occurrence of Y varies under different interventions on X, then there is a causal relationship between X and Y. If one analyzes precisely, beyond intuitive vagueness, it becomes evident that this expression is incomplete and omits much information. First, the limitations of '$P(X)$' in situations involving continuous variables have already been analyzed in Section 2. Secondly, this type of expression lacks information about other conditions. Without tacit understanding, an uninformed person might interpret it as suggesting that "the relationship between X and Y can be inferred by comparing $P(Y|do(X))$ and $P(Y|do(\neg X))$, regardless of the similarities and differences of other conditions." Of course, such an inference could be incorrect.

In understanding Theorem 5, it is important to note that these are not formed through empirical analysis and summarization but are derived through a quasi-axiomatic approach.

Therefore, they are fundamentally metaphysical. However, in situations where information for reasoning is incomplete, epistemological knowledge may help enhance the credibility of causal inferences. This epistemological information might include spatiotemporal relationships, results of intervention, background knowledge, knowledge about microscopic structure (mechanisms), etc. And this is exactly how the elements of traditional causal theories truly contribute to causal reasoning.

7 Proper use of series concepts related to causation

Now, we can provide a comprehensive clarification on the connotation of causation. From an ontological level, the *if-then* function is the basis of the relationship in PDD, which is the basis of the concept of causation in the PDD framework.

Based on Theorem 3, there are two types of contexts involving the concept of causation: one emphasizes the potential influence of conditions on outcomes in the general sense, while the other focuses on attributing responsibility for the variation of the state of affairs in specific cases.

In the general sense context, the definition of general potential influential relation from the epistemological perspective is given in Theorem 5:

$$S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$$

The standard statement for this relationship is: "conditional on Ψ, $S(X)$ can influence $S(Y)$," or "$S(X)$ is a potential factor influencing $S(Y)$." When the background condition Ψ is stable, it might be suitable to say that "there is a *causal relationship* between $S(X)$ and $S(Y)$", or "there is a *causal relationship* between X and Y." Due to the possibility of changing background conditions, it is inappropriate to use concepts of 'cause' and 'effect' in this sense, which may be done in traditional causal theories.

In the context of specific events, the identification of the factors responsible for the events

from the epistemological perspective ultimately relies on a combination of Theorem 5 and Theorem 4[7]:

$$\exists \Delta S(X), \quad S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$$

Therefore, the standard statement for specific events is: "under the specific condition Ψ, $\Delta S(X)$ is the *cause* of $\Delta S(Y)$" , or "under the specific condition Ψ,$\Delta S(Y)$ is the *effect* of $\Delta S(X)$". The subjects referred to as cause and effect are the variations of the state of affairs $\Delta S(X)$ and $\Delta S(Y)$, not the specific state $S_i(X)$ or the variable X or specific value $X = x_i$, the latter is believed by other theories of causation. Additionally, meaningful causal statements must specify background information Ψ, while the everyday expressions in this context are actually simplified and left this information as part of tacit knowledge.

The above encompasses almost the entirety of the concept of causation in the new theory presented in this paper. This deviation in expression may also be one of the important reasons why the concept of causation has been unclear historically.

8 Comparison with RCT and interventionism

This section compares the new theory presented in this paper with several currently popular theories of causation, and further argues for the reasonableness and completeness of the new theory.

Whether in the statement of scientific knowledge or in theories of causation, a fundamental consensus is the necessity of a *ceteris paribus* clause, literally meaning "other things being equal". However, for a long time, its detailed implications were seldom clearly defined. In recent years, some causal theories have advanced in this area and achieved some basic consensus. According to these theories (referred to as C-theories here), when analyzing the

[7]When multiple factors contribute to a specific result, the composite explanation ultimately reduces to the identification of each individual factor, which still follows the method in Theorem 5.

causal relationship between two objects, all the scenarios that need to be considered can be roughly expressed as in Figure 4. Here, the analysis focuses on the causal relationship between X and Y; A might influence X, and is often termed as 'confounder' when it also independently influences Y; M is influenced by X and affects Y, typically called 'mediator'; U independently affects Y; and N represents things unrelated to both X and Y. Specifically, C-theory asserts that the ideal environment for causal reasoning requires the same quantities of A and U in two comparable scenarios, but it does not necessitate the same amount of M (Cartwright, 1979; Woodward, 2016a; Pearl and Mackenzie, 2018).

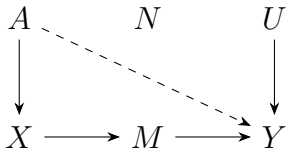


Figure 4: Typical situations involved in the definition of causation

This approach has significant intuitive appeal. Firstly, it confines all conditions that could potentially influence Y apart from X, naturally leading to the attribution of the variation of Y to the variation of X. Secondly, in observational data, if M were confined, Y would not change anymore, thereby obstructing the judgment of the relationship between X and Y.

However, C-theory also has some flaws. Firstly, it typically only briefly asserts that the quantity of variable M need not be the same, which is a description at the level of *type*. This can pose problems in cases where, suppose, there is a certain amount of M pre-existing in both comparative scenarios, and their quantities differ. Since a part of M independently influences Y, the difference in Y cannot be reliably attributed entirely to the difference in X.

Secondly, C-theory typically does not impose restrictions on unrelated variable N. All things required to be constrained are causally connected with X or Y. For example, in Pearl's structural equation representation $Y = f(X, U_Y)$, U_Y represents omitted things that *affect* Y (Pearl and Mackenzie, 2018, p. 276). However, if one does not yet know what the meaning of causation is, how can these objects be defined? In other words, C-theory's

definition of the causal relationship between X and Y depends on prior knowledge involving the concept of causation, which is a flaw for definition and reasoning without prior causal knowledge.

In contrast, Theorem 5 not only fully encompasses the content intended to be defined but also lacks the above flaws. In equation (24), the identical $\mathcal{E},\mathcal{N},\mathcal{I}$ on both sides of the equation represent all things required to be constrained. $\mathcal{E}$ includes all factors that can influence Y, covering U in Figure 4; $\mathcal{I}$ covers general conditions that can influence $S(X)$, thus including A in Figure 4. Hence, A and U, which C-theory wishes to keep constant, are similarly constrained in Theorem 5.

Regarding the mediator variable M, the analysis of conditions is conducted at the individual level in equation (24), not at the type level. The pre-existing part of M in the environment is considered as part of $\mathcal{N}$. Since $\mathcal{N}$ is required to be constant, its definition implies that the pre-existing part of M must remain constant, while other parts of M are not restricted.

Regarding the unrelated variable N, in equation (24), it falls into $\mathcal{N}$ and is required to be the same in both comparison scenarios. Moreover, it is undifferentiated from other causal information. This approach ensures that the defined causal relationship fundamentally does not depend on prior knowledge involving causation.

Therefore, the theory presented in this paper not only meets the goals set out by C-theory for defining causal relationships but also improves upon them. The sameness of Ψ (the combination of $\mathcal{E},\mathcal{N},\mathcal{I}$) in both comparison scenarios provides a clear and rational interpretation for the *ceteris paribus* clause.

A successful causal theory naturally cannot avoid comparison with the method of Randomized Controlled Trials (RCTs), which is widely regarded by the scientific community as an ideal method for analyzing causal relationships. The basic idea of an RCT hinges on the random assignments of participants into experimental and control groups, ensuring, theoretically, an equal distribution of factors that could influence the results across these groups.

By comparing results from the treatment group, who receive the intervention, with the control group, who receive a placebo or no intervention, RCTs aim to attribute differences in outcomes directly to the intervention.

Comparing RCTs with Theorem 5, it becomes apparent that the reasoning conditions created by RCTs almost perfectly align with the elements required by Theorem 5. Random assignment ensures that all "other conditions are equal" between the experimental and control groups, which is precisely the same as expressed by Ψ (the whole of $\mathcal{E}, \mathcal{N}, \mathcal{I}$) and Theorem 5. This means that both related and unrelated conditions, undifferentiated, are balanced between the two groups, also largely independent of prior knowledge. The difference in intervention between the two groups signifies the difference between $S_1(X)$ and $S_2(X)$. The logic of inference in RCTs and Theorem 5 are also similar: when the above conditions are met, the difference in outcomes between the two scenarios can be attributed to the difference between $S_1(X)$ and $S_2(X)$. In short, RCTs can be seen as artificially creating scenarios that fit Theorem 5.

Thus, theoretically, the explanatory range of Theorem 5 perfectly encompasses that of RCTs. However, RCTs cannot serve as a definition for the more general concept of causation due to its clear limitations in applicability. In scenarios where RCTs are infeasible due to ethical or practical reasons, they cannot be relied upon to explain relationships between objects. For example, understanding the interactions between celestial bodies clearly cannot involve RCTs. In contrast, the definition in Theorem 5 of this paper does not have this limitation, as it is a metaphysical definition.

This naturally leads to the question of the role of '*intervention*' in the concept of causation. In the realm of causal theory, the concept of intervention is extensively discussed. Holland (1986) emphasizes the necessity of interventions in drawing causal inferences in randomized experiments. Rubin's Causal Model posits that interventions are critical in defining causal effects, distinguishing between potential outcomes under intervention and no intervention scenarios (Rubin, 2005). Pearl introduces the '*do*-operator' (e.g., $do(X)$) as a formal

representation of intervention (Pearl and Mackenzie, 2018, p. 49). Woodward elaborates on this by defining an intervention as an idealized manipulation that changes only the factor under consideration, thereby clarifying its causal influence on an outcome (Woodward, 2003, 2016a).

However, the framework for causal explanations based on interventions still has imperfections. These extend beyond scenarios such as celestial interactions, where interventions cannot be implemented, and include the following aspects.

Firstly, interventionism contains a certain degree of circularity in its complete definition of causation. It relies on the analysis of results similar to Figure 4, where interventions are used to fix variables X, A, and U. Interventionism aims to define what a causal relationship is (between X and Y), but it requires prior knowledge of causation to define the relationship between M and X and Y. In other words, its explanation of the concept of causation relies on the concept of causation itself, making it non-reductive. Woodward argues that this is not viciously circular, because it is not explaining the same causal relationship with itself but using causal knowledge of one relationship (M and X) to explain another causal relationship (X and Y), and some basic hypothesis of the former can provide new insights for the latter (Woodward, 2016a). However, this defense faces a fundamental difficulty: if one does not understand what causation means, they will first fail to comprehend the relationship between X and M, and thus, also the relationship between X and Y.

Secondly, conclusions about causal relationships based on interventionism may not be complete and accurate. Consider a thought experiment depicted in Figure 5, where there is a room A with a switch S_A connected to a screen D_B in another room B, displaying the action signal (ON/OFF) of S_A. There is another switch S_B in room B, actually connected to the light L_A in room A, and a person P_B in room B may press S_B according to her will upon seeing the screen D_B. Suppose a person P_A enters room A and wants to determine if switch S_A controls light L_A.

If P_A repeatedly operates S_A and observes L_A lighting up, can he deduce the relationship

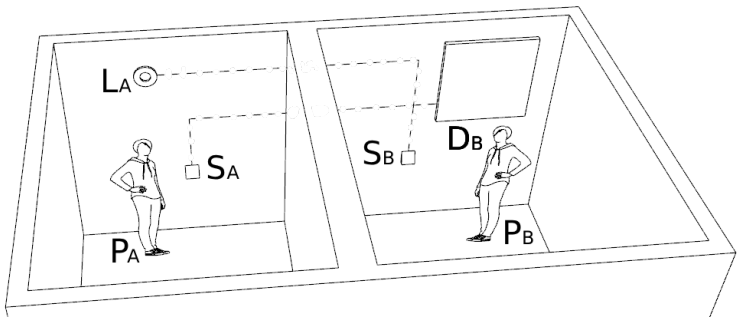


Figure 5: Switch and light thought experiment.

between the two through this intervention experiment? According to interventionist theory, if P_A finds that L_A changes following repeated interventions (operating S_A), they might conclude that S_A can stably control L_A. This case challenges interventionist causal theory by demonstrating that interventionist reasoning can lead to erroneous attributions of cause, especially in complex or hidden causal mechanisms, and in situations where not all relevant factors can be controlled or observed. According to the new theory in this paper, P_B's will is part of the condition Ψ, but it's unclear if this condition remains consistent between P_A operating the switch and not. Therefore, it does not completely satisfy the judgment condition in Theorem 5, preventing a strict inference of causation.

Thirdly, it is possible to possess prior causal knowledge without referring to intervention. For example, in the eyes of a car designer, the causal relationships between various mechanical parts of a car (such as 'accelerator' and 'wheels') are constructed based on mechanical principles. This causal knowledge is a priori for the designer and does not depend on intervention to be known. Even if the designer is physically unable to test it personally due to a disability, as long as their knowledge of mechanisms is sound and they fully trust their design skills, they can possess this causal knowledge. More metaphysically, even if a specific car component fails, this does not affect the causal relationships between components on the design blueprint, which is a metaphysical piece of knowledge independent of any real conditions. For a computer programmer, the influence relationship between events in the program is defined by him, and he can obviously know this relationship without any

intervention. Even more extremely, suppose the world was created by an omniscient God; in the eyes of this God, the relationships between states of affairs in the world would be a priori known and not dependent on intervention.

Therefore, intervention may not be a necessary concept for defining causation and may not always help in accurately discerning causal relationships at the epistemological level. Thus, a comprehensive definition of causation cannot rely on the concept of intervention.

9 Cases application

Randomized Controlled Trials (RCTs) are widely regarded in the scientific community as the ideal method for analyzing causal relationships. As previously analyzed, under circumstances where intervention is possible, the new theory of causation presented in this paper is fundamentally aligned with RCTs in its definition of causal relationships. However, in some situations, RCTs are not feasible due to ethical or practical reasons. In this section, several cases are presented to demonstrate how the new causal theory can provide insights into causal analysis in such scenarios.

Snow (1856) conducted an extensive investigation into the cholera outbreaks in London in the mid-19th century. Within the areas he studied, two main water companies served a population exceeding 300,000. He found that the mortality rate was 107 per 100,000 inhabitants supplied by one company, compared to only 8 per 100,000 for the other. His investigation revealed that the infrastructure of these companies extended throughout the districts, with pipes running down all streets and into nearly every court and alley. These companies supplied water indiscriminately to a diverse demographic, encompassing both affluent and impoverished residents, large and small houses, and people of all genders and ages. Notably, there was no distinction in the socioeconomic status or occupation of individuals receiving water from either company.

Is it possible to infer a causal relationship based solely on this information, using the

new causal theory presented in this paper, without conducting an interventional trial? The answer is yes.

We can apply the information from Snow's study as follows: Use $S(X)$ to represent the state of the water supply (either from one company or the other), and use $S(Y)$ to represent the incidence of cholera (quantified as mortality rates). Since factors like socioeconomic status, age, gender, and occupation, consistent across populations served by both companies are almost the same (all those things which can coexist with both $S_1(X)$ and $S_2(X)$ are the same in both populations), they can be represented as Ψ. Thus, all the information satisfy $S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$. According to Theorem 5, we can confidently infer that $\Delta S(X)$ is the cause of $\Delta S(Y)$, which means the difference of water supply is cause of the difference of the cholera mortality rates (107 vs. 8 per 100,000).

Therefore, for this case, based on the new theory of causation presented in this paper, a causal relationship can be confidently inferred from the aforementioned non-interventional observational data. The inference of this relationship is quite substantial. Further investigation into this causal relationship is not necessary for the judgment of the causal relationship, but only deepen our understanding of the details of this relationship, thereby strengthening our recognition of the relevant causal mechanisms. For example, Snow's further investigations revealed that the key difference between the two water companies lay in their water sources. One company drew water from contaminated sections of the lower Thames River, while the other's water source was relatively cleaner, as it was sourced from the upper reaches of the Thames. This established a clear relationship between water source contamination and cholera mortality rates, and made our conclusion more strong.

The new theory can also provide ideas for exploring the relationship between smoking and lung cancer.The historical debate on the link between smoking and lung cancer hinged on the difficulty of proving causation without randomized controlled trials. Skeptics argued that the observed association might be due to unmeasured confounding factors, such as a genetic predisposition influencing both smoking habits and lung cancer risk. This skepticism,

underscored by the fact that not all smokers develop lung cancer and some non-smokers do. The practical and ethical challenges in conducting RCTs for smoking meant that these doubts persisted for decades (Pearl and Mackenzie, 2018, pp. 168-169).

Firstly, we need to precisely state the question of interest, which is the relationship between tobacco inhalation (rather than the act of smoking itself) and lung cancer. Let $S(X)$ represent the state of tobacco inhalation and $S(Y)$ represent lung health status. According to Theorem 5, an effective judgment environment requires that all conditions potentially coexisting with both $S_1(X)$(non-inhalation of tobacco) and $S_2(X)$(inhalation of tobacco) remain the same in both comparison scenarios. These conditions include all possible internal body conditions related to smoking, such as a hypothesized smoking gene, because a person with a smoking gene may or may not inhale tobacco. Therefore, data should not be derived from active smokers; instead, it should be obtained from passive smokers, who are exposed to environments where smoking occurs by others, thereby inhaling tobacco smoke.

From this perspective, focusing on passive smokers, two methods are feasible for investigating the relationship between tobacco and lung cancer. The first method can resort to traditional randomized controlled trial without facing ethical issues. For instance, we could intervene in the environment of passive smokers to reduce their inhalation of tobacco smoke, and then compare the differences in disease incidence rates between them and an uninterfered group, as was done by Emmons et al. (2001).

The second method employs a statistical investigation method similar to the cholera case mentioned above, we can construct two comparative groups: one group consisting of passive smokers and another group not exposed to passive smoking. During the investigation, strategies can be employed to ensure that the two groups are as similar as possible in various aspects such as gender, age, income, education level, and all others that are potentially coexisting with both $S_1(X)$and $S_2(X)$. Such meticulous matching will ensure that Ψ remain consistent across both groups. Consequently, according to Theorem 5, if the investigation reveals that the lung cancer incidence rate is significantly higher in the passive smoking

group compared to the non-passive smoking group, $S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$, then it would strongly suggest a causal relationship between tobacco smoke inhalation and an increased risk of lung cancer.

In this case, utilizing the new causal theory, we can clearly determine that passive smokers should be the focus of the analysis. This approach transforms scenarios initially unsuitable for RCTs into feasible ones. Additionally, a statistical investigation strategy can construct conditions required for determining causal relationships, enabling inference without necessarily depending on interventional trials. This offers a highly beneficial analytical tool for research methods in social science and epidemiology.

10 Conclusion

The concept of causation is essential for understanding relationships among various phenomena, yet its fundamental nature and the criteria for establishing it continue to be debated. This paper presents a new theory of causation through an axiomatic-like system. The core of this framework is *Probability Distribution Determinism* (PDD), which updates traditional determinism by representing states of affairs as probability distributions, with the '*if... then...*' function serving as its foundational definition. Based on PDD, by merely using appropriate naming strategies, it is possible to derive systems in which the structural characteristics of relationships among things closely resemble those in the real world, such as having various forms of nested hierarchies. Additionally, there are two interrelated yet distinctly different contexts associated with relationships in PDD: one emphasizes the potential influence of conditions on outcomes in the general sense, while the other focuses on attributing responsibility for the state changes in specific scenarios. The formula for determining the relationship in the general sense is established as $S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$. Subsequently, within the PDD framework, the paper clarifies the legitimate use of a series of concepts related to causation in those two contexts, thus encompassing the entire detailed connotation of the

concept of causation. The comparison with other theories of causation and the analysis of case applications demonstrate that the new theory is applicable not only to situations where other theories are competent but also to situations where they are not. This suggests that, although certain aspects within the new framework may require further analysis, it provides a highly promising direction for a deeper understanding of causation.

This paper presents a new theory of causation through an axiomatic-like system.It proposes that states can be represented as probability distributions, an approach more broadly applicable in causal analysis than traditional concrete state understanding and probability interpretation. Traditional determinism, combined with this new understanding, can accommodate both non-probabilistic and probabilistic scenarios. This paper introduces it as probabilistic distribution determinism, with its core being the logical conditional if...then.... Based on this new assumption of determinism, a system can be derived where the structural characteristics of the relationships between objects closely resemble the real world, supporting the reasonableness of the assumption.

The framework leads to the analysis that there are two interrelated yet distinct contexts regarding the relationships between states: one focuses on the potential influence of conditions on results in an absolute sense, and the other on the ascription of responsibility for changes in states in specific contexts. The cause as the bearer of explanation in the latter must be something that has a state difference in that context, not something that is relatively stable and unchanged. The traditional usage of the concept of causation conflates these two different levels, and distinguishing them helps to clarify the concept's meaning.

Within the framework of probabilistic distribution determinism, the paper derives the essence of causation as: in the absolute sense context, the definition of general relation is, $S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$, with the standard statement for this general potential influence relationship being: conditional on Ψ, $S(X)$ can influence $S(Y)$ or $S(X)$ is a potential factor influencing $S(Y)$. At this level, due to the possibility of changing background conditions, it is inappropriate to use cause and effect to describe entities. If the background

condition Ψ is stable, then a causal relationship between $S(X)$ and $S(Y)$, or X and Y, can be affirmed. In the context of specific events, for the identification of the responsible party for a particular change in state, there is a variation in the state of $S(X)$ in the current context, and$S(Y|S_1(X), \Psi) \not\equiv S(Y|S_2(X), \Psi)$. Thus, the standard causal statement for specific events is: conditional on Ψ, $\Delta S(X)$ is the cause of $\Delta S(Y)$ (or $\Delta S(Y)$ is the effect of $\Delta S(X)$), meaning causal statements must specify conditional information, and the subjects referred to as cause and effect are the state variations ($\Delta S(X)$ and $\Delta S(Y)$), not the absolute states $S_i(X)$ or the variable X.

Through comparison with RCTs and interventionism, the paper demonstrates that the new theory's expression through Ψ fully includes what these theories aim to control when defining causal relationships, and is more precise in its definition, avoiding their limitations. Case analysis shows that the new theory can be applied to infer causal relationships from observational data not based on experimental intervention and provide insights for scientific inquiry, thereby uncovering more scenarios where interventional trials are feasible.

In summary, existing causal theories do not fully suffice in defining the concept of causation, whereas the explanatory framework of this paper not only explains scenarios applicable to other causal theories but also those where other theories are not applicable. Its powerful explanatory capacity is demonstrated through comparison with traditional theories and analysis of specific cases. While some details of the analysis and arguments in the paper may need further development and strengthening, the framework provided holds potential for resolving disputes related to the concept of causation.

References

Cartwright, N. (1979). Causal Laws and Effective Strategies. *Noûs 13*(4), 419–437.

Cartwright, N. (2004). Causation: One word, many things. *Philosophy of Science 71*(5), 805–819.

DeGroot, M. H. and M. J. Schervish (2012). *Probability and Statistics* (4th ed.). Boston: Addison-Wesley.

Emmons, K. M., S. K. Hammond, J. L. Fava, W. F. Velicer, J. L. Evans, and A. D. Monroe (2001). A randomized trial to reduce passive smoke exposure in low-income households with young children. *Pediatrics 108*(1), 18–24.

Ford, K. W. (2004). *The Quantum World: Quantum Physics for Everyone*. Cambridge, Massachusetts: Harvard University Press.

Gallow, J. D. (2016). A theory of structural determination. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition 173*(1), 159–186.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis 44*(1), 49–71.

Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science 69*(S3), S342–S353.

Glennan, S. (2010). Mechanisms, Causes, and the Layered Model of the World. *Philosophy and Phenomenological Research 81*(2), 362–381.

Hitchcock, C. R. (1993). A generalized probabilistic theory of causal relevance. *Synthese 97*(3), 335–364.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association 81*(396), 945–960.

Hume, D. (2007). *An Enquiry Concerning Human Understanding*. Oxford World's Classics. Oxford: Oxford University Press.

Kuhn, T. S. (2012). *The Structure of Scientific Revolutions* (4th ed.). Chicago: The University of Chicago Press.

Lewis, D. (1973). Causation. *The Journal of Philosophy 70*(17), 556–567.

Machamer, P., L. Darden, and C. F. Craver (2000). Thinking about Mechanisms. *Philosophy of Science 67*(1), 1–25.

McDonnell, N. (2018). Transitivity and proportionality in causation. *Synthese 195*(3), 1211–1229.

Menzies, P. and H. Beebee (2020). Counterfactual Theories of Causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020 ed.). Metaphysics Research Lab, Stanford University.

Mumford, S. and R. L. Anjum (2013). *Causation: A Very Short Introduction* (1st ed.). Number 371 in Very Short Introduction. Oxford: Oxford University Press.

Pearl, J. and D. Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect* (1st ed.). New York: Basic Books.

Rubin, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association 100*(469), 322–331.

Russell, B. (1913). I.—On the Notion of Cause. *Proceedings of the Aristotelian Society 13*(1), 1–26.

Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis (2017). Mastering the game of Go without human knowledge. *Nature 550*(7676), 354–359.

Snow, J. (1856). Cholera and the Water Supply in the South Districts of London in 1854. *Journal of Public Health, and Sanitary Review 2*(7), 239–257.

Wheeler, B. (2022). Causation in a Virtual World: A Mechanistic Approach. *Philosophy & Technology 35*(1), 7.

Wittgenstein, L. (2009). *Philosophical Investigations* (4th ed.). Chichester: Wiley-Blackwell.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford Studies in Philosophy of Science. New York: Oxford University Press.

Woodward, J. (2016a). Causation and Manipulability. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Metaphysics Research Lab, Stanford University.

Woodward, J. (2016b). The problem of variable choice. *Synthese 193*(4), 1047–1072.