Article

Philosophy of the Social Sciences
2022, Vol. 52(1-2) 3–24

DOI: 10.1177/00483931211049766
journals.sagepub.com/home/pos
SAGE

Interventionism and Over-Time Causal Analysis in Social Sciences

Tung-Ying Wu[1]

Abstract

The interventionist theory of causation has been advertised as an empirically informed and more nuanced approach to causality than the competing theories. However, previous literature has not yet analyzed the regression discontinuity (hereafter, RD) and the difference-in-differences (hereafter, DD) within an interventionist framework. In this paper, I point out several drawbacks of using the interventionist methodology for justifying the DD and RD designs. Nevertheless, I argue that the first step toward enhancing our understanding of the DD and RD designs from an interventionist perspective is to take advantage of the assumptions of common trend and continuity.

Keywords

Methodology, interventionism, causation, regression discontinuity, difference-in-differences

1. Introduction

The interventionist theory of causation (hereafter, interventionism) has been advertised as an empirically informed and more nuanced approach to causality than the competing theories. Specifically, interventionism can be methodologically

[1]Institute of Philosophy, Chinese Academy of Sciences (CASIP), China

Corresponding Author:
Tung-Ying Wu, Institute of Philosophy, Chinese Academy of Sciences, No.19A Yuquan Road, Beijing 100049, China.
Email: tungyingwu@outlook.com

useful and illuminating to associate causal claims with what the results of hypothetical experiments would be without doing the experiment in the way described by interventionism. The notion of a hypothetical experiment is primarily useful in understanding non-experimental design and can figure as a regulative and clarifying ideal in causal inference (Woodward 2007, 163; Woodward 2015, 3587). This is especially the case in those areas of the social sciences in which it is difficult to carry out actual experiments. Several attempts have been made to relate quantitative causal analysis in social sciences within an interventionist framework (Greene 2020; Moneta and Russo 2014; Runhardt 2015; Russo 2011, 2012, 2014; Woodward 2003, 2007, 2015). Particularly, some argue that interventionism justifies the methodology of causal discovery in social sciences by claiming that the use of instrumental variables is readily justifiable based on interventionism (Woodward 2015, 3593–94).[1] To the best of my knowledge, however, previous literature has not yet explored other methods and principles for social research, such as the regression discontinuity (hereafter, RD) design and the difference-in-differences (hereafter, DD) design from an interventionist perspective.[2] How does one characterize the DD and RD designs within the interventionist framework? Is interventionism an appropriate methodology for assessing and justifying the DD and RD designs? Is it methodologically useful to construe causal claims made with the DD and RD methods as a hypothetical intervention in an ideal experiment?

By examining the above questions, I found it is difficult to provide the interventionist justification of the DD and RD designs. One of the significant difficulties of characterizing these methods within the interventionist framework is that the DD and RD compare outcomes at different times. Since they let outcomes evolve with time, such dynamic evolution of outcomes might confound experimental results and lead to biased estimations of causal effects. Therefore, from the interventionist perspective, the RD and DD are

[1]Some writers have raised several issues about applying interventionism to causal analysis in social sciences. For instance, structural equations models might aim to unveil or characterize the underlying mechanism, not just invariance under counterfactual interventions (Moneta and Russo 2014; Russo 2012). Sometimes scientists just cannot intervene on a variable and interventionists cannot provide any test in such circumstances (Kowalenko 2017; Russo 2014; Wunsch et al. 2014). The interventionist criteria of variable choice is difficult to apply in the social sciences (Greene 2020). Interventionism fails to assess the validity of the whole model (Russo 2012). The goal of the social sciences is to find lawful relationships (Boumans 2003). Interventionism is too restrictive as it "imposes" that causal relations are tested by one specific method (Cartwright 2007). Policy interventions are consequences of causal relations previously established (Russo 2014). The fragility of causal structure in social contexts may break down even under the most surgical intervention (Reiss 2009; Steel 2007). However, previous writers have not dealt with how to understand the DD and RD methods from the interventionist perspective in much detail.

[2]Woodward briefly mentions that the RD can be justified from the interventionist perspective (Woodward 2015, 3593). However, as I will indicate below, the interventionist justification of RD is problematic.

"bad" methods of causal analysis by the methodological standards described by interventionism because these methods do not control for (or "held fixed") all other covariates or factors that are associated with the passage of time and may mask the causal effects of interest.

I propose several ideas for interventionism to circumvent the difficulties presented above. In brief, the assumptions of common trend and continuity behind the DD and RD methods enable the interventionist methodology to provide a comprehensive explanation and justification for the DD and RD. Also, these assumptions block confounding causal channels via uncontrolled time-related factors, even if these channels are intrinsic to the DD and RD methods. I argue that the DD and RD methods in social research are justifiable based on interventionism and can lead to a better understanding of the DD and RD within the interventionist framework.

I organize the paper into six sections. Section 2 gives a brief review of interventionism and several motivations for its methodology. Section 3 and 4 will consider the methods of DD and RD, respectively, and will include the discussions of the difficulties preventing the DD and RD from the interventionist construal. Section 5 brings together the lessons from the previous section and offers some improvements that support the interventionist justification of the DD and RD designs. Section 6 is the conclusion.

2. Interventionism

Woodward develops interventionism from the theory of causal Bayes nets and the analysis of intervention from the causal modeling literature to characterize the (type-level) causal efficacy of variables on other variables (Woodward 2003).[3,4] Interventionism aims to non-reductively illuminate the nature of causal relations by characterizing the interconnections of various causal concepts such as correlation, intervention, counterfactual, and causal explanation (Woodward 2015, 3581–82, 3585–87). The idea is that under some specific circumstances, variable X causes variable Y if and only if manipulating X leads to changes in Y. In detail, Woodward provides the following conditions of causation and intervention:

[3]The theory of causal Bayes nets can formally represent causal relations in a rigorous mathematical language. Not only can it represent causal relationships among variables of interest but identify the causal effect of an intervention. For the purpose of this paper, I will not introduce causal Bayes nets in detail. Suffice it to say its strength lies in its rules to infer causal relations from a causal model's probabilistic distributions and the correspondent causal graph with its axioms. I will expand on the notion of intervention later.

[4]Variables (denoted by uppercase letters) represent tokens of events or properties that serve as relata of (type-level) causal relations, and these variables range over possible values that represent these events' occurrence or non-occurrence or properties' instantiation or non-instantiation, or a value if an event or property is of a quantity.

(M) A necessary and sufficient condition for X to be a cause of Y with respect to a set of variables V is that there exists a possible intervention on X that will change Y when one holds fixed all other variables in V that are not on the direct path between X and Y (Woodward 2003, 59).[5]

Simply put, X's causal effect on Y can be empirically tested by examining whether manipulating X makes a difference in Y while controlling for (or holding fixed) all other variables in V. Note that (M) appeals to the notion of intervention and its definition is (IV):

(IV) I is an intervention for X with respect to Y if and only if (1) I causes X; (2) when I obtains certain values, then X ceases to depend on other variables except I; (3) I causally influences other variables only through X; (4) I is independent of any cause of Y that is on a directed path that does not go through X (Woodward 2003, 98).

An intervention externally sets X to a specific value and makes it no longer depend on its original parents, and the intervention neither causes any variable other than X nor is caused by any other variable in a causal model (Woodward 2003, 95–98).[6] To elaborate, in a traditional randomized experiment, a randomized device determines the assignment of treatment to control for any possible common cause of treatment and outcomes. For example, a researcher is uncertain whether high-carb diets or a kind of DNA causes obesity. She observes that one who has a high-carb diet is likely to develop obesity but suspects that it is the obesity-induced DNA that causes one to have a high-carb diet and obesity. How does she determine the causal relations between a high-carb diet and obesity? Scientists exploit randomized experimental designs to solve the problem. They use a randomized device to determine whether a subject receives a high-carb diet (called treatment) or not, and the researcher can compare those who received a high-carb diet (called the treatment group) and those who did not (called the control group). If subjects in the treatment group are more likely than the control group to develop obesity, the researcher can safely conclude that a high-carb diet causes obesity. The benefit of using the randomized device is that it disables all pre-existing causes of a high-fat diet, so the DNA no longer exerts its influence on one's diet. The notion of intervention follows precisely the rationale behind randomized experiments. An intervention described by (IV) directly changes X. It disables all pre-existing causes of X so a research design that uses interventions can distinguish causation from the correlation that muddles our judgment about what causes what, thus reach a more accurate evaluation of causal effects.

[5] V is a set that contains variables whose causal relationships we are interested in studying. "M" stands for "manipulation theory." Note that (M) is not intended as a reductive analysis of causal relations. As emphasized above, interventionism aims to non-reductively illuminate the nature of causal relations by interpreting a causal claim as a claim about the outcome of a hypothetical experiment where an intervention on a cause changes its effect (Woodward 2015, 3593–94).

[6] "X's parents" stand for X's immediate causes.

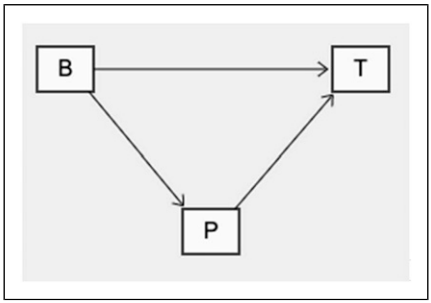

Figure 1. Hesslow's example.

(IV) and (M) seem to capture the methods of causal analysis in social sciences. In empirical studies, this is done by randomized trials as previously mentioned. In a non-experimental setting where social scientists cannot perform a randomized experiment (called quasi-experiments), however, they may use data to identify specific causal relations as if the data result from randomized trials under certain conditions.[7] (M) and (IV) naturally suggest these conditions. For instance, if one conceives a policy as a hypothetical intervention that satisfies (IV), and the data shows that a change of *X* leads to a change of *Y*, then one can infer a causal relation between *X* and *Y* as warranted by (M) (Woodward 2015, 3593–94). Either way, the methods of causal inference in experimental or non-experimental settings seem to exemplify interventionist ideas.

One crucial element of (M) is its requirement to control for all other variables while examining what the result of a hypothetical intervention would be. This strict requirement is motivated by Hesslow's example. Suppose that taking birth control pills (*B*) increases the probability of thrombosis (*T*), that is, *B* directly causes *T* with a certain probability.[8] However, birth control pills might prevent pregnancy (*P*), and pregnancy is likely to cause thrombosis as well. So, *B* decreases the probability of *P,* and *P* increases the probability of *T.* Hence, *B* might directly give rise to *T.* However, B might also deter *T* because *B* inhibits *P* that engenders *T.* As it happens, *B*'s direct positive probabilistic

[7]In a quasi-experiment if one can reasonably assume that there is no common cause of the assignment of treatment and outcomes, then one may treat the data as if it is the result of a randomized experiment conducted by nature (a hypothetical intervention) (Angrist and Pischke 2009, 21; Morgan and Winship [2007] 2014, 9).

[8]"Direct causal relation" means that a cause does not cause its effect by causing other events or variables and there is no intermediate property between the cause and its effect. In other words, the causal influence a cause exerts on its effect is not mediated by any other property. "X direct causes Y" means that X does not cause Y by causing other properties, namely, there is no intermediate property between X and Y.

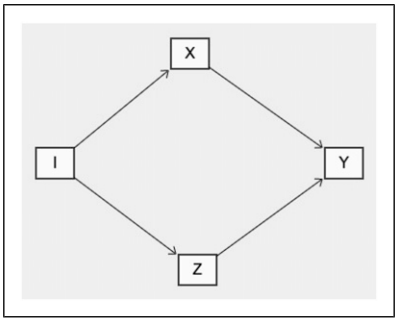

Figure 2. An intervention causes X and Z, and both X and Z cause Y.

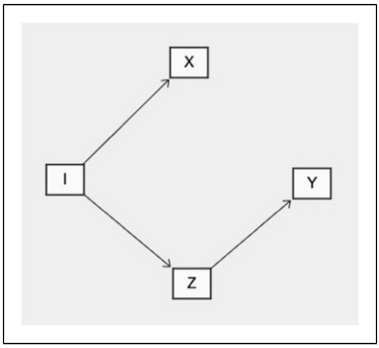

Figure 3. An intervention causes X and Z, but only Z causes Y.

efficacy on *T* is canceled out by *B*'s negative probabilistic influence on *T* by preventing *P*. Consequently, if one takes a birth control pill, the net change of probability of *T* is 0 (Hesslow 1976; Woodward 2003, 49–50). Figure 1 is the causal diagram of Hesslow's example.[9]

In Figure 1, *B* exerts its positive causal effect on *T* along the route from *B* to *T*, but *B* also impedes the occurrence of *T* by preventing *P*, which is a cause of *T*. Since, in balance, *B*'s positive and negative influence on *T* is canceled out, manipulating *B* does not lead to a probabilistic change of *T*, and falsely conclude that *B* does not cause *T*.

To avoid cases where a causal relation is muddled by other factors, (M) requires that one assesses the result of a hypothetical intervention while holding fixed all other variables. For instance, if one controls for *P* by only observing subjects who are not pregnant and manipulates B by introducing birth control

[9]A causal diagram is a directed acyclic graph that consists of nodes that represent variables in V, and arrows between nodes that represent causal relations. If the value of a variable Y depends on X, then there will be a directed path from X to Y.

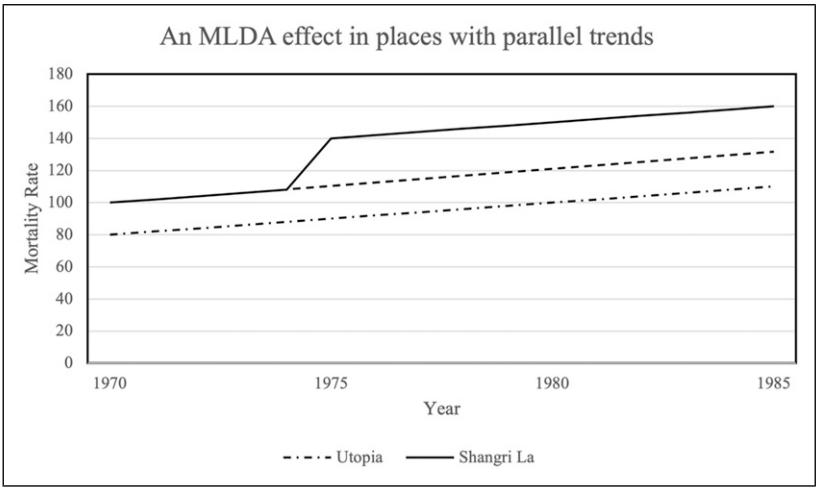

Figure 4. The relationship between mortality rate (death rate per 100,000) and year in Shangri La and Utopia.

pills, the probability of *T* will increase so that one can confirm the causal relation between *B* and *T*. Therefore, (M) stipulates that to discern a causal relationship between cause and effect one must control for all other variables.

Turning now to (IV). It is important to note that (IV) demands that an intervention causally influences an outcome only through *X*. If an intervention causally affects an outcome through *X* and other variables, one simply cannot out rule out the possibility that the change of the outcome is entirely due to other variables rather than *X*. For example, the following Figure 2 shows that an intervention causes *X* and *Z*, and both *X* and *Z* cause *Y* (Woodward 2003, 102–103).

Now, suppose that when the intervention *I* changes *X,* one observes the change of *Y*. However, if there is no causal relation between *X* and *Y*, one might also observe the change of *Y* while intervening on *X*. Hence, one cannot discern whether the actual causal structure is Figure 2 or Figure 3. In Figure 3, even if the intervention *I* changes *X*, one might still observe the change of *Y*. It might be the case that the change of *Y* is entirely due to *Z* rather than *X*. One cannot determine that it is *X* that does the causal job. Perhaps, it is *Z* that does all the causal job. If an intervention causally influences *Y* through variables other than *X,* then it might lead to a false judgment that *X* causes *Y*. Therefore, to rule out the possibility of indiscernible causal structures, (IV) strictly requires that a qualified intervention must causally influence *Y* only through *X* (Woodward 2003, 102).[10]

[10]This requirement is called the exclusion restriction in social research: a single causal channel connects instruments with outcomes (Angrist and Pischke 2009, 116–17).

However, as I will argue below, interventionism cannot account for DD and RD that are common techniques for social scientists to make causal inferences in quasi-experiments. Before explaining these problems, I will now move on to introduce the DD and RD designs.

3. Interventionism and Differences-In-Differences Design

Many research questions in social sciences are inquiries about causality. For example, does earning a college degree increase future job income? Does a policy of reducing the minimum legal drinking age (hereafter, MLDA) decrease the mortality rate of young adults? These questions are about whether X causes Y or how significant the causal effect of X on Y is. In non-experiment settings, social scientists have developed various techniques for the causal analysis of statistical data.[11] When analyzing longitudinal data, the DD design is a non-experimental technique that is adequate for making causal inferences (Angrist and Pischke 2009, 227–43; Morgan and Winship [2007] 2014, 359–60).

Longitudinal data contain repeated observations of the same individuals over a period, so social scientists refer to them as panel data or time-series data (Morgan and Winship [2007] 2014, 363–64). For example, researchers who study the causal effect of MLDA have access to samples of young adults observed over multiple years before and after they are 21 years old. With some assumptions, time-series data may be used for estimating causal effects, and the power of over-time observation is considerable (Morgan and Winship [2007] 2014, 354–55).

To demonstrate DD, suppose that in a study of the causal effect of reducing the MLDA to 18 years old in 1975 (hereafter, the treatment) when the original MLDA is 21 years old, a researcher obtains the data that comprises repeated observations of two places—Shangri La and its neighbor Utopia—before and after 1975. Shangri La received the treatment and but not Utopia, in which the MLDA maintains 21 years old. Also, one divides the data into two periods: the first period is from pretreatment time, and the second period is from post-treatment time. Figure 4 plots the relationship between mortality rate (death rate per 100,000) and year in Shangri La and Utopia.[12]

[11]These techniques include regression analysis, structural equations models, and instrumental variables.

[12]I draw this example with modifications from Angrist and Pischke (2014, 198). I take these models or examples as epistemic artifacts that make a theoretical point and explain alternative solutions and inferences. In particular, Knuuttila's (2009, 222–24) analysis of economics models seems reasonable. She concludes that the aim of unrealistic or fictional economics models is to understand and draw inferences from using or manipulating them as they help us think clearly and proceed systematically. Instead of constructing a complex model that accurately corresponds to its target, the goal of unrealistic models or examples is to help us learn and try out different frameworks of inferences from building and manipulating them.

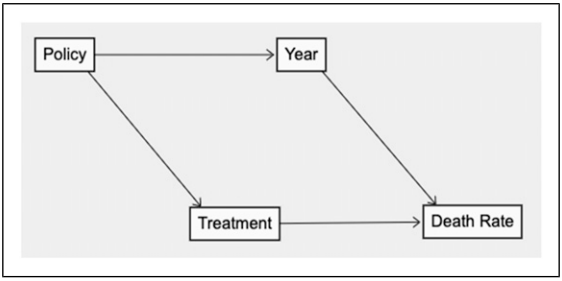

Figure 5. The causal diagram of the example in Figure 4.

The most interesting aspect of this figure is that the death rates in both places move in parallel before 1975, but when Shangri La reduces its MLDA, its death rate jumped in 1975.[13,14] The dotted line represents what the trend of the mortality rate would have been in Shangri La had all developed as it did in Utopia. Another significant aspect is that the death rates in both places always increase with time. Suppose that the constant increase in death rates in both places is for reasons unrelated to the policy. For instance, there was a pandemic outbreak from 1970 to 1985 or an ongoing war with foreign countries. Further, suppose that both places are comparable, so there is no significant difference between Shangri La and Utopia. Hence, one may take Shangri La as the treatment group and Utopia as the control group.

The most reasonable explanation for the jump in 1975 is that Shangri La's reducing MLDA caused the jump. Since the trend of the death rate in Shangri La is parallel to Utopia before and immediately after 1975, a fair expectation is that Shangri La and Utopia should have moved in parallel during 1975. However, instead of remaining parallel, the death rate of Shangri La suddenly increased around 1975, and the parallel resumed soon after 1975. In addition, the significant difference between the two places in 1975 is that Shangri La reduced its MLDA but not Utopia. Therefore, the unexpectedly sudden jump of Shangri La in 1975 is undoubtedly due to the treatment.

The next question is how the DD matches what interventionism conceives as a form of inference about the outcomes of hypothetical experiments. How does interventionism as a methodological ideal evaluate the DD design?

[13]These lines are derived by the regression analysis, which attempts to determine a best-fitting linear approximation of the relationship between two quantities (in this example, death rate and year). I assume the relationship between the death rate and year is liner.

[14]Even if the two trends of the death rates in both places do not move in exact parallel a more technical approach such as parametric regression models can be adopted to control for the difference in trends. (Angrist and Pischke 2014, 197; Morgan and Winship [2007] 2014, 368). For the sake of simplicity, I assume that they move in parallel. I will elaborate more on this assumption in Section 5.

Presumably, interventionism requires that the policy is a qualified intervention satisfying (IV) and the causal inference made by the DD method satisfies (M). However, as I will argue below, the DD design does not comply with (IV) and (M), so it is not justifiable on the ground of interventionism.

To begin with, the policy of reducing MLDA satisfies the first two conditions of (IV). The policy externally causes the treatment to rule out any potential common cause between the treatment and its outcome. Suppose that which place (Shangri La or Utopia) one lives in is for reasons unrelated to the policy of MLDA, so the assignment of treatment to each subject is as if randomly assigned by nature.[15] It seems plausible to conclude that the policy makes the legal access to alcohol no longer dependent on its original causes, even if there is a common cause of the legal access to alcohol and death rates.[16] Hence, the policy meets the first two requirements of (IV).

However, the DD method still does not provide a suitable intervention in the example. Recall that the third condition of (IV) requires that an intervention causally influences an outcome only through its cause. Nonetheless, the policy does not causally influence the outcome only through legal access to alcohol. It has a separate causal channel that influences Shangri La's mortality not via the legal access to alcohol. To illustrate, as shown in Figure 4, the slopes of death rates are constantly positive. This means that some other factors keep influencing the death rates (called the time effect). More importantly, the time effect is different each year. In 1970, the time effect on the mortality in Shangri La is 100, but in 1974 the time effect is (roughly) 110. Accordingly, the policy not only reduces the MLDA but determines which year to place it. Therefore, the policy does not meet the third condition of (IV) and cannot be a qualified intervention. The causal diagram of the example is Figure 5:

As discussed in Section 2, the policy does not satisfy (IV) because of the worry that the time effect ("year" in the above diagram) might do all the causal work. It might just happen that the time effect in 1975 tremendously raises the mortality rate. Within the framework of interventionism, there is no room for the policy to be an intervention. Though interventionists seem to disregard

[15]See footnote 7.

[16]For example, before placing the policy, whether one has the legal access to alcohol depends on local bar tenders and these bar tenders also causally influence the death rates. Clearly, the attitude of bar tenders is a common cause that confounds one's judgment of the causal relation between access to alcohol and death. It might be that a bar tender that allows the access to alcohol for teenagers below 21 tend to be more short-lived (more death) because they care less about health. So, it might be that the correlation between access to alcohol and death is entirely due to a bar tender. In contrast, the government's policy of setting MLDA disables bar tenders' causal influence on the access to alcohol. Hence, by intervening on MLDA via the policy one can empirically verify the causal claim that teenagers' access to alcohol causes the change of morality rate if the data shows a correlation between them.

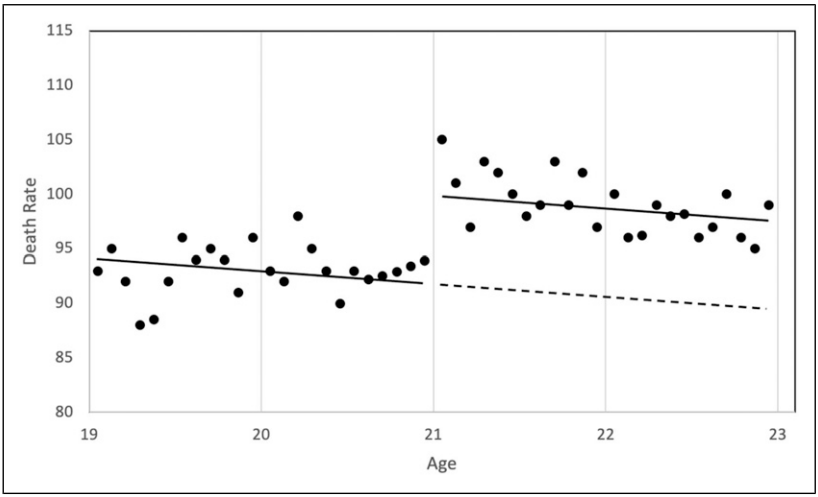

Figure 6. The relationship between the death rates and the ages in Shangri La.

"policy" in Figure 5 as a qualified intervention, social scientists have no problem using it to make the causal inference.

Here is how social scientists use the DD design to infer causal relations: compare the difference between the divergence of each group's outcome before and after treatment. Specifically, the DD method isolates the treatment effect by using the control group to isolate the effects generated by factors other than the treatment (Angrist and Pischke 2009, 229; Angrist and Pischke 2014, 194; Morgan and Winship [2007] 2014, 364–73). For example, let t_1 be 1970 and t_2 be 1980. The difference of Shangri La's outcomes at t_1 and t_2 is $150-100 = 50$. The difference of Utopia's outcomes at t_1 and t_2 is $100-80 = 20$. Hence, the empirical evidence for the causal relationship between treatment and outcome is the difference between the difference of Shangri La's outcomes at t_1 and t_2 and the difference of Utopia's outcomes at t_1 and t_2: $50-20 = 30$. Equivalently, one may obtain the same result by measuring the divergence between Shangri La's outcome in t_2 and the dotted line's outcome at t_2. Take the year 1980, for example. Shangri La's death rate is 150, and the death rate by the dotted line is 120, so the empirical evidence for the causal relation is $150-120 = 30$.

Though one might estimate the causal effect of a treatment by simply comparing the treatment group before and after the treatment, the estimation might pick up the time effects of other factors that increase the death rate every year. For example, if one estimates the causal effect of the treatment by the divergence between Shangri La's death rate in 1972 (105) and its death rate in 1980 (150), one will overestimate the causal effect ($150-105 = 45$). Hence,

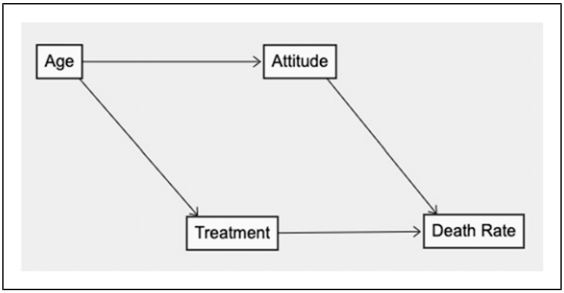

Figure 7. The causal diagram of this example in Figure 6.

the DD method accurately estimates the causal effect by eliminating the time effect that influenced observed outcomes around the time of treatment.

Here is the second difficulty for interventionism to accommodate the DD design. Recall that according to (M), for X to be a cause of Y with respect to a set of variables V is that there exists a possible intervention on X that will change Y when one holds fixed all other variables in V. However, a common strategy for the DD method to determine a causal relationship is simply to compare the difference of each group's outcome before and after treatment.[17] Instead of testing a causal relation while controlling for all other variables as required by (M), the DD's procedure for causal inference does not control for the time variable. It lets time effects run its course and compare subjects at two different points in time. Hence, the DD's method for causal inference is not correct since it defies the interventionist normative standards of causal inference. What seems to be an impossible procedure for causal inference from the interventionist perspective is actually practiced by social scientists.

Having discussed the difficulties that interventionism cannot make sense of the DD design, I will now move on to present the RD design briefly and argue that interventionism also has similar difficulties explaining it.

4. Interventionism and Regression Discontinuity Design

Let us now look at the RD design. It is another quasi-experimental design that undertakes a causal analysis as if in a randomized controlled experiment. Social scientists apply it when data shows that a trend is shifted around the treatment and the treatment assignment is sharply discontinuous.[18] Moreover, the RD design requires sufficient observations and information about subjects

[17]Technically speaking, the difference of each group's outcome before and after treatment must be statistically significant (comparing to the standard error) to establish a causal relation. Otherwise, the difference might be entirely attributed to the sampling variance or random factors other than treatment effect (Angrist and Pischke 2009, 230).

[18]I will explain this more below.

around a cutoff.[19] Though it does not work for all causal inquiries, RD has been popular among social scientists in many quasi-experiments (Angrist and Pischke 2009, 251–59; Morgan and Winship [2007] 2014, 360–63).

I will use the previous example to demonstrate the RD. For the sake of simplicity, suppose that Shangri La places a policy of MLDA (the treatment in this example): an individual whose age is above 21 gets legal access to alcohol, but no one whose age is below 21 gets legal access to alcohol.[20] Also, only the data from Shangri La is available. The research question is whether the policy causes the change in the death rate. Figure 6 represents the relationship between the death rates and the ages of the same group of subjects:[21]

This figure shows the death rate in a population aged 19–23, and those dots represent death rates (calculated as the number of deaths per 1 million persons per year) by month of age (defined as 30 days intervals). For example, the second dot (from the left) between ages 19 and 20 stands for the death rate for individuals whose age is in the interval between 230-month-old and 231-month-old. The regression analysis derives the two solid lines that show the best-fitting linear approximation of the relationship between death rates and age in months.[22] They stand for the trends of death rates before and after the treatment from the data. The dotted line stands for the counterfactual of what the mortality rate would have been had those individuals below 21 had not received the treatment (i.e., the legal access to alcohol). The most significant feature in Figure 6 is that the slopes of the death rate are negative. Though these dots in Figure 6 fluctuate between 90 and 100 before the treatment and 95 and 105 after the treatment, the regression analysis shows these two lines continue to trend downward (Angrist and Pischke 2014, 149–50). As discussed in the previous example, there must be other factors that constantly decrease young adults' mortality rate. Suppose that these subjects care more about their safety and health when they grow older. Subjects of different ages have different mortality rates because of their attitudes that vary with age.

The RD evaluates the data in Figure 6 as if it is a randomized experiment conducted by nature. There is no common cause of treatment assignment (whoever is above 21 gets legal access to alcohol) and the death rate. Moreover, whether a subject is above or below 21 is random. Therefore, in this

[19]A "cutoff" is a boundary between subjects who received treatment and those who do not. In Figure 5 below, the cutoff is 21 years old.

[20]The RD method requires the assumption that no one below 21 (called "cutoff") gets the treatment and everybody above the cutoff gets the treatment. It is in this sense that the assignment of treatment is discontinuous. If the assumption is unsubstantiated, social scientists sometimes adopt the "fuzzy" RD method to solve the problem (Angrist and Pischke 2009, 259–67; Morgan and Winship [2007] 2014, 362).

[21]I draw this example with modifications from Angrist and Pischke (2014, 150).

[22]I assume the relationship between death rates and age in month is liner.

example, it is as if nature randomly assigns treatment to subjects living in Shangri La (Angrist and Pischke 2009, 259).

Since the RD design does not control time-related variables, the RD resists the interventionist construal. What prevents the RD from the interventionist reading are variations of the previous difficulties of explaining the DD. The first obstacle is that the age variable is not an intervention under (IV). The second is that the RD's procedure for inferring causal relations is not justifiable based on (M).

Here is the first obstacle. The age variable does not satisfy (IV) because it has a separate causal channel that influences mortality not via the treatment. To see this, note that subjects with different ages and stages of maturity react differently to their safety and health, so their death rates vary with age. As a result, Shangri La's policy of selecting the drinking age also picks up the potential effects of their attitudes toward their lives. Therefore, the age variable has two causal channels that influence mortality: one is the legal access to alcohol and the other is the attitude that varies with maturity. This explains why the policy is not a qualified intervention under (IV), and the RD is not suitable for the interventionist construal. The causal diagram of this example is Figure 7

Let us consider how the RD infers the causal relation by measuring the divergence between the solid line to the right and the dotted line.[23] Here is the rationale for the RD design. Since no control group is available, the RD method does not use a control group to answer what the trend would have been if Shangri La had not received the treatment. However, one way to answer the counterfactual question is to suppose that the future resembles the past. Namely, those just under 21 are a good comparison for those just above 21. Hence, one may use those under 21 as the control group for those above 21 to predict what would have happened if the treatment group had not had legal access to alcohol. If the assumption is reasonable, one can use the trend of the mortality rate of those under 21 to infer the trend of those above 21 and draw the dotted line. Hence, the difference between the trend of those above 21 and the trend of the dotted line represents the evidence for the causal effect generated by the policy of MLDA. Consequently, the RD method compares the same group exposed to the treatment state and the control state at different points in time (Angrist and Pischke 2009, 252; Morgan and Winship [2007] 2014, 361).

The RD's procedure for causal inference also introduces the second obstacle for interventionism. Recall that from the interventionist perspective, the procedure for establishing a causal relation is that a hypothetical intervention on X changes Y while holding fixed all other variables according to (M). However, in the example, since one does not have the control group, the fundamental nature of causal inference of the RD design is that one does not observe the subjects with the same age in both the treatment state and the control state. Instead, the RD method compares death rates for people above

[23] The divergence must be statistically significant, that is, large enough to evidentially support the existence of a causal relation. See footnote 17.

and below 21, so it is impossible to hold age fixed in the RD design. Thus, (M) cannot justify the RD's procedure for inferring causal relations.

Taken together, I have identified two variations of problems that challenge the interventionist evaluation of the causal analysis in social sciences by examining the DD and RD designs. First, since the variables used as interventions correlate with the time-related effects on the death rates, they are not proper interventions based on the interventionist conception of causation. Second, the practices of inferring causal relations in the DD and RD do not follow the interventionist methodology that requires holding all other variables fixed. As a result, if one evaluates the DD and RD based on the methodology advised by interventionism, the DD and RD would be flawed or unreliable. This seems to be an implausible implication since they are common approaches in causal analysis in social research. Evidently, the interventionist methodology of causal inference is not suitable for characterizing them and is not representative of causal analysis in social sciences.

However, it might be too early to dismiss the interventionist project. Interestingly, its methodology offers the characteristics of an ideal randomized experiment, and these characteristics might deepen our understanding of those non-experimental techniques of causal analysis in social sciences from a different perspective. It seems that the problems that I raised above are the reasons that the interventionist methodology needs revisions, and there is certainly room for improvements. In the next section, I will present several such improvements.

5. Revisions and Improvements

This section begins by addressing the problem that the DD's and RD's strategies for causal inference call into question the generality of (M). It will then engage with the problem that the DD and RD do not follow the standard of (IV).

Let us now consider the DD method of identifying causal relations. As argued in Section 2, Hesslow's example in Figure 1 motivates the requirement that testing a causal relation involves holding all variables fixed other than cause and effect, but the DD does not control for the time variable. For the DD, one way to ease the difficulty is to adopt the common trend assumption: treatment and control groups would have exhibited the same trend without the treatment (Angrist and Pischke 2009, 230; Angrist and Pischke 2014, 185). The common trend assumption supports the dotted line in Figure 4 and justifies the DD's strategy of obtaining causal relations by measuring the difference between the line of Shangri La's death rate after 1975 and the dotted line.[24] Without the

[24]Recall that the dotted line that stands for what would have happened had Shangri La had not reduced its MLDA. As I indicated above, the DD design infers causal relations by comparing the difference between the divergence of each group's outcome before and after treatment. This procedure is equivalent to comparing the difference between the line of Shangri La's death rate after 1975 and the dotted line in Figure 4.

common trend assumption, it is almost impossible to infer the causal relationship between the policy and the death rate.[25] For example, it might be the case that without the treatment, Shangri La's death rate after 1975 is exactly 30 more than it was before 1975, so the policy did not cause any change in the death rate. Thus, the common trend assumption is crucial to the DD method.

Though (M) prescribes that the conditions for the policy to be a cause of the death rate is that a hypothetical intervention on the policy changes the death rate while strictly holding the time variable fixed, I suggest a broader understanding of this requirement. The requirement (M) will be more comprehensive if it adds that under the common trend assumption, one can test causal relations by holding a fixed temporal interval instead of a fixed point in time. Specifically, when comparing each group's outcome before and after treatment, the time of observations before and after treatment must be fixed at a specific value for each group. Given the common trend assumption, (M) will be more inclusive if it admits the substitution of a temporal interval for a point in time given the common trend assumption.

To demonstrate, let Y be outcomes, subscript tg1 be treatment group at time t_1, subscript cg1 be control group at time t_1, time t_1 be a specific time before treatment, and t_2 be a specific time after treatment. $(Y_{tg2} - Y_{tg1}) - (Y_{cg2} - Y_{cg1})$ is the formula for how the DD method captures the causal difference (Angrist and Pischke 2009, 229; Angrist and Pischke 2014, 184; Morgan and Winship [2007] 2014, 364, 372).[26] Now, the more comprehensive understanding I suggested is that even if the DD procedure does not hold fixed the time variable, it at least holds the time interval (t_1 to t_2) fixed. In the absence of a fixed time interval, the time effects on both groups will be different, and it is impossible to arrive at a very reliable estimate of causal differences.[27] In contrast, under the common trend assumption, a fixed time interval for treatment and control groups implies the same time effects on them. Hence, assessing a causal relation while holding a fixed time interval is equivalent to assessing the causal relation while holding a fixed time effect. Thus, the strategy that I just devised eliminates the concern from Hesslow's example that other factors might bias one's judgment of causal relations and might be helpful for interventionism in solving the difficulty of justifying the DD method.

Turning now to the RD method, the rationale of RD's causal inference relies on the continuity assumption that all other unobserved determinants of an outcome are continuous at cutoff except treatment assignment (Angrist and Pischke 2009, 252–53; Angrist and Pischke 2014, 153; Morgan and Winship [2007] 2014, 356). For example, in Figure 6, the continuity assumption implies that those dots

[25] See footnote 14.

[26] The formula stands for the DD's procedure of comparing the difference between the difference of each group's outcome before and after treatment.

[27] For example, suppose that one controls for different time intervals for treatment and control groups and captures causal differences by $(Y_{tg1985} - Y_{tg1970}) - (Y_{cg1980} - Y_{cg1975})$. It will, of course, overestimate the causal difference generated by the policy.

(subjects) near the cutoff (the vertical line that stands for 21 years old) are comparable. As discussed above, the fundamental nature of the RD method is that it takes a group of subjects below the cutoff as a control group for those above the cutoff. However, the problem is that different ages come with different attitudes toward risk. Those who received the treatment (i.e., getting access to alcohol) are older than those who did not receive the treatment, and the resulting change of death rate might be entirely attributed to the difference of their age instead of the treatment. For example, it might just happen that young adults on their 21st birthday, and on that day only, prefer to engage with risky activities, and the jump of death rate can be entirely attributed to that attitude of risk-seeking. However, the continuity assumption implies that subjects in the neighborhood around the cutoff are continuous. All other factors that influence outcomes of those subjects (including the age-related factors such as attitude) are similar in a significant way if they are adjacent to the cutoff. Therefore, given the continuity assumption, all other determinants of outcomes of those subjects who are below and above 21 years old, including their ages and attitudes, are assumed to be approximately the same.[28]

As for (M), it requires that one condition for a causal relation strictly requires holding the time-related variable fixed. I suggest that with the aid of the continuity assumption, one way to make the interventionist proposal (M) more valuable for our understanding of the RD design is to add that one may treat those subjects around age 21 as if they are at the same age given the continuity assumption. The problem is that the RD inference's nature does not control for the age variable. However, under the RD's continuity assumption, it is acceptable to relax the requirement of strictly holding fixed the age variable at precisely 21 and allow the age variable to take a value within a narrow range centered around age 21. Thus, if one tests a causal relation while holding time-related variables within a narrow range of values around the cutoff, it will be equivalent to verifying the causal relation while holding fixed the influences

[28]Note that I assume no hidden cause other than the policy and the time effect (risk attitude that varies with age) for the outcome in Figure 6. One possible worry is that it might just happen that those young adults on their 21st birthday, and on that day only, prefer to engage with risky activities (so at the cutoff point the curve would rise). Or it might just happen that those young adults on their 21st birthday prefer to avoid any risk (so at the cutoff point the curve would flatten out). However, as I indicated above, the rationale of RD's causal inference essentially relies on the continuity assumption to estimate what would have happened after the cutoff point. The continuity assumption indicates that all other unobserved determinants of an outcome are continuous at cutoff except treatment assignment. Therefore, in Figure 6, the continuity assumption implies that subjects in the neighborhood around the cutoff are assumed to be approximately the same, including their ages and attitudes toward risk. In other words, their attitudes toward risk on their 21st birthday are not radically different from their attitude before and after their 21st birthday. This assumption enables one to use those subjects under 21 to predict what would have happened if the treatment group had not had legal access to alcohol. In particular, one can use the trend of the mortality rate of those under 21 to infer the trend of the mortality rate of those above 21 and draw the dotted line. I thank an anonymous referee for pressing me to clarify this point.

generated by these time-related variables. The suggestion I just offered eases the worry that other factors might bias one's judgment of causal relations.

Having solved the problem with (M), I will now discuss the problem with (IV) that requires that an intervention can only have one causal channel from the intervention to an outcome. If my suggestions above are plausible to some extent, controlling for a time interval or a narrow range of values around the cutoff appears to block those muddling causal channels from the intervention to an outcome. As discussed in the previous section, these unwanted causal channels, such as the path from policy to death rate via year and the path from age to death rate via attitude, create unintended biases that should be considered. However, the improvements I suggested above rule out these biases by holding fixed time effects, and the observed causal differences could be safely attributed to treatments. Hence, though one worries that we cannot disentangle time effect from treatment effect, in the DD and RD designs, the assumptions of common trend and continuity enable one to isolate time effects from treatment effects. Thus, it is methodologically fruitful to relax the requirement that an intervention must influence an outcome only through treatment, given the common trend and continuity assumptions.[29,30]

[29]Another possible response to the problems I raised in Section 5 is to appeal to the regression method. It might be argued that the regression method can control for variables such as "year" and "attitude." However, the RD and DD designs are best seen as distinct tools that differ importantly from the regression method. First, note that in common practices of the regression method, we compare treatment and control outcomes at specific values of control variables, in the hope that treatment is as good as randomly assigned after conditioning on controls. Nevertheless, in Figure 6 there is no value of the age variable at which we get to observe both treatment and control state at the same time. Second, the DD design uses a control group to isolate treatment effect from time effect but there is no such procedure in the regression equations. Third, some argue that regression equations are merely reduced form equations. They do not represent true causal structure and carry counterfactual information about possible interventions. Fourth, the RD's and DD's regression equations do not represent distinct equations that correctly and completely describe the results of hypothetical interventions on all the endogenous variables in the representation. They simply aim for estimating coefficient or statistical information. It seems that these equations do not take the interventionist reading about what the results of possible interventions would be while holding fixed all other variables. See also footnote 1. For reasons of space, I will not expand on the regression equations of the RD and DD designs.

[30]One question that needs to be raised is that the assumptions of common trend and continuity might themselves generate problems that require additional theories and assumptions, but then a regress of correction or justification looms. These two assumptions might create problems in two possible ways. First, these assumptions might be problematic in some cases where one needs additional assumptions to correct these problems. Take Figure 4 for instance. Suppose that the constant increase of the death rate in Shangri La is due to a pandemic, but in Utopia, it results from an ongoing war. Since Utopia might agree to a cease-fire, but Shangri La failed to keep the pandemic under control, the death rates in both places might not move in parallel after 1975. In addition, it is also possible that even if the causes for the constant increase of the death rates in both places are the same, and no significant difference between these places has been found. However, as a result of

6. Conclusion

The interventionist theory of causation uses a causal modeling framework based on intervention analysis to characterize the causal efficacy of properties on other properties. It identifies causal relations by predicting what would have happened under a hypothetical intervention under some specific conditions. Specifically, interventionism appears to make causal analysis in social sciences determinate, clear, and precise by associating causal claims with hypothetical experiments in the way described by (M) (Woodward 2015, 3589).

However, due care must be paid in applying the interventionist methodology to the context of causal analysis that appears in the social sciences literature. This paper questions the interventionist justification of some methods and principles of causal inference in social research, especially the DD design that is a technique of over-time causal analysis and the RD design that is structurally similar to over-time causal analysis. One might think that the RD and DD designs are nothing but the potential outcome models, which belong to a methodology based on interventions, and presume that the interventionist account is a suitable approach for methodological assessment in those contexts in social research (Woodward 2015, 3585). However, I

some factors unbeknownst to us (for example, aside from the policy of MLDA, a hidden and deadly event might happen in 1975 in Shangri La), the death rates in both places would not have moved in parallel after 1975 and the common trend assumption is just misleading and problematic. These examples show that the common trend assumption generates new problems in some cases and these cases require additional assumptions to solve these new problems. More generally, if treatment and control groups are different in some respect, or there is a hidden cause that is not observable to all concerned, the assumptions of common trend and continuity will be problematic. Therefore, in Figures 4 and 6, I assume no significant difference between treatment and control groups and no hidden cause. Nevertheless, one might doubt that one will always find such problematic cases for these assumptions, and no sufficient number of additional assumptions will stop these problematic cases from arising. Still, it does not seem unreasonable to say that one should use a process to arrive at a balance between one's principles and judgments about these particular cases. However, an in-depth evaluation of the process would go beyond the scope of this paper. Second, these two assumptions might generate new problems because they depend on additional assumptions and theories. Specifically, the assumptions of commend trend and continuity depend on a background theory about how the death rate of a population in a place evolves with time. Without the background theory that justifies these assumptions, these unjustified assumptions will render the methods of DD and RD implausible. Therefore, the commend trend and continuity assumptions require additional assumptions and background theories. However, these additional assumptions and theories themselves need further assumptions and theories to justify them. One might wonder whether they lead to a regress of justifications in a way that is related to the problem of induction. Still, some commentators suggest that the regress seems less severe, and the chain of assumptions and theories might simply terminate in what appears to be conventional grounds or brute facts (Okasha 2005, 251–53, Norton 2003, 668). A complete discussion of this problem lies beyond the scope of this paper. I thank an anonymous referee of this journal for raising this point.

argue that the interventionist methodology does not explain why the DD's and RD's causal inferences do not control for time-related variables. I also point out some considerations in these non-experimental designs that help the interventionist account have something to say relevant to the methodological assessment of these DD and RD designs. I submit that the first step toward enhancing our understanding of the RD and DD designs within the framework of interventionism is to take advantage of the RD's and DD's assumptions of common trend and continuity. Indeed, finding a causal link between cause and effect while letting time runs its course is highly problematic, but this might depend on the methodology and assumptions chosen for one's research.

Acknowledgments

I would like to thank Chuang Liu for his continuous support and guidance. Thanks also to the two anonymous reviewers of this journal and my longtime friends at the reading group: "Causal Reasoning and Causal Models," for providing me with the learning opportunities that was essential in helping me complete this paper. I also received helpful feedback from the audiences at the 2021 Joint Meeting of the Asian Network for the Philosophy of the Social Sciences, the European Network for the Philosophy of the Social Sciences, and the Philosophy of Social Science Roundtable (Hitotsubashi University, Tokyo, Japan. Online).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Angrist, J. D., and Pischke, J.-S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Angrist, J. D., and Pischke, J.-S. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.

Boumans, M. 2003. "How to Design Galilean Fall Experiments in Economics." *Philosophy of Science* 70 (2): 308-29.

Cartwright, N. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics*. Cambridge, UK: Cambridge University Press.

Greene, C. 2020. "Nomadic Concepts, Variable Choice, and the Social Sciences." *Philosophy of the Social Sciences* 50 (1): 3-22. DOI: 10.1177/0048393119878783

Hesslow, G. 1976. "Two Notes on the Probabilistic Approach to Causality." *Philosophy of Science* 43 (2): 290–92.

Knuuttila, T. 2009. "Representation, Idealization, and Fiction in Economics: From the Assumptions Issue to the Epistemology of Modeling." In *Fictions in Science: Philosophical Essays on Modeling and Idealization*, edited by E. Suárez, 205–31. New York: Routledge.

Kowalenko, R. 2017. "Manipulationism, *Ceteris Paribus* Laws, and the Bugbear of Background Knowledge." *International Studies in the Philosophy of Science* 31 (3): 261-83. DOI: 10.1080/02698595.2018.1463693

Moneta, A., and Russo, F. 2014 "Causal Models and Evidential Pluralism in Econometrics." *Journal of Economic Methodology* 21 (1): 54-76. DOI: 10.1080/1350178X.2014.886473

Morgan, S. L., and Winship, C. (2007) 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. 2nd ed. New York: Cambridge University Press. DOI: 10.1017/CBO9781107587991

Norton, J. D. 2003. "A Material Theory of Induction." *Philosophy of Science* 70 (4): 647-70. DOI: 10.1086/378858

Okasha, S. 2005. "Does Hume's Argument Against Induction Rest on a Quantifier-Shift Fallacy?" *Proceedings of the Aristotelian Society* 105 (1): 237-55. DOI: 10.1111/j.0066-7373.2004.00113.x.

Reiss, J. 2009. "Causation in the Social Sciences: Evidence, Inference, and Purpose." *Philosophy of the Social Sciences* 39 (1): 20-40. DOI: 10.1177/0048393108328150.

Runhardt, R. W. 2015. "Evidence for Causal Mechanisms in Social Science: Recommendations from Woodward's Manipulability Theory of Causation." *Philosophy of Science* 82 (5): 1296-1307. DOI: 10.1086/683679.

Russo, F. 2011. "Correlational Data, Causal Hypotheses, and Validity." *Journal for General Philosophy of Science* 42 (1): 85-107. DOI: 10.1007/s10838-011-9157-x.

Russo, F. 2012. "On Empirical Generalisations." In Probabilities, Laws, and Structures, edited by D. Dieks, W. J. González, S. Hartmann, M. Stöltzner, and M. Weber, 123-39. Dordrecht: Springer. DOI: 10.1007/978-94-007-3030-4_9

Russo, F. 2014. "What Invariance Is and How to Test for It." *International Studies in the Philosophy of Science* 28 (2): 157-83. DOI: 10.1080/02698595.2014.932528

Steel, D. 2007. *Across the Boundaries: Extrapolation in Biology and Social Science*. New York: Oxford University Press.

Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Woodward, J. 2007. "Causal Models in the Social Sciences." In Philosophy of Anthropology and Sociology, edited by S. P. Turner, and M. W. Risjord, 157-210. Amsterdam: Elsevier. DOI: 10.1016/B978-044451542-1/50006-4

Woodward, J. 2015. "Methodology, Ontology, and Interventionism." *Synthese* 192 (11): 3577-99. DOI: 10.1007/s11229-014-0479-1

Wunsch, G., Mouchart, M., and Russo, F. 2014. "Functions and Mechanisms in Structural-Modelling Explanations." *Journal for General Philosophy of Science* 45 (1): 187-208.

Author Biography

Tung-Ying Wu is an assistant professor at the Institute of Philosophy, Chinese Academy of Sciences (CASIP) in Beijing. After working as a legal practitioner for several years, he decided to pursue philosophy for his interest in philosophical significance. He earned a Master of Arts degree in philosophy from University of Missouri-St. Louis and received a doctorate from University of Missouri-Columbia on the topic of decision theory and causal reasoning. Before working at CASIP, he was a postdoctoral researcher at Fudan University in Shanghai, where he began his exploration of the philosophy of causality and causal modeling.