Stochastic Mechanistic Interaction

Carlo Berzuini*   A. Philip Dawid†

June 17, 2014

Abstract

We propose a fully probabilistic formulation of the notion of mechanistic interaction (interaction in some fundamental mechanistic sense) between the effects of putative (possibly continuous) causal factors $A$ and $B$ on a binary outcome variable $Y$. We define mechanistic interaction in terms of departure from a generalized “noisy OR” model, under which the multiplicative causal effect of $A$ (resp., $B$) on the probability of failure ($Y = 1$) cannot be enhanced by manipulating $B$ (resp., $A$). We present conditions under which mechanistic interaction in the above sense can be assessed via simple tests on excess risk or superadditivity, in a possibly retrospective regime of observation. These conditions are defined in terms of conditional independence relationships that can often be checked on a graphical representation of the problem. Inference about mechanistic interaction between direct, or path-specific, causal effects can be accommodated in the proposed framework. The method is illustrated with the aid of a study in experimental psychology.

Keywords: Biological mechanism, causal inference, compositional epistasis, direct effects, directed acyclic graphs, excess risk, experimental psychology, independent effects, noisy OR, observational studies, path-specific effects, superadditivity.

*Centre for Biostatistics, Institute of Population Health, The University of Manchester, Manchester, UK
†Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, UK
1 Introduction

Consider an outcome variable $Y$ that responds to manipulations of two variables, $A$ and $B$. We are interested in whether the effects of $A$ and $B$ on $Y$ interact in some fundamental mechanistic sense. For example, we might be interested in whether an environmental exposure $A$ interferes with the effect of a drug $B$ on a disease $Y$ at some mechanistic—presumably molecular—level. Such a relationship, which we shall make more formal in a later section of this paper, we call mechanistic interaction.

One might begin to investigate mechanistic interaction by fitting a regression model of the dependence of $Y$ on $(A, B)$ and then testing for presence of statistical $A \times B$ interaction, but such a test will depend on the chosen response scale, and will generally not be interpretable in any deep mechanistic sense. Hence the need for a mathematical formalization of mechanistic (as opposed to statistical) interaction, and of the conditions under which this phenomenon can be detected from empirical data via appropriate, response-scale independent, statistical tests. In many applications, discovery of mechanistic interaction could represent a step forward in the understanding of the studied system. In genetics, evidence of mechanistic interaction between two genes with respect to a phenotype of interest could point to the molecular mechanisms implicated (Bernardinelli et al., 2012).

Ideally we would wish to assess mechanistic interaction by a controlled experiment, but this is often not possible or not convenient. Various authors have proposed tests for inferring mechanistic interaction (suitably defined) from observational data (Rothman, 1976; Rothman & Greenland, 1998; Greenland & Poole, 1988; VanderWeele & Robins, 2008, 2009; VanderWeele, 2009, 2010, 2011; VanderWeele & Laird, 2011). Consider, for example, the case that $A, B$ and $Y$ are all binary variables, with $Y = 1$ referred to as “failure” and $Y = 0$ as “survival”, and let $C$ denote a further (possibly empty) set of observed variables. Let $R_{abc}$ denote the observational risk of failure, $Y = 1$, conditional on $A = a, B = b, C = c$. Then, in certain observational situations, and under certain conditions, the following properties (of which the first is stronger than the second) have been shown to imply some form of mechanistic interaction:

**Excess risk:**

\[
R_{11c} - R_{10c} - R_{01c} > 0. \quad (1)
\]
Superadditivity:

\[ R_{11c} - R_{10c} - R_{01c} + R_{00c} > 0. \]  \hspace{1cm} (2)

These can be alternatively expressed as:

Excess risk:

\[ S_{10c} + S_{01c} - S_{11c} > 1, \]  \hspace{1cm} (3)

Superadditivity:

\[ S_{10c} + S_{01c} - S_{11c} - S_{00c} > 0, \]  \hspace{1cm} (4)

where \( S_{ijc} := 1 - R_{ijc} \) is the corresponding probability of survival (of \( Y = 0 \)). An important property of the above tests is that they are (at least approximately, under assumptions) testable under retrospective sampling. More precisely, the above conditions give criteria for synergistic mechanistic interaction between \( A \) and \( B \) in producing \( Y \), in that the combined effect (suitably measured) of increases in \( A \) and in \( B \) to increase \( Y \) is greater than expected on the basis of their individual effects. This is the interpretation we shall maintain here. The case of antagonistic mechanistic interaction, where the the combined effect is smaller than expected, is readily handled by interchanging the values 0 and 1 for \( Y \), and interchanging the \( R \)s and \( S \)s in equations (1)–(4) and elsewhere.

Most work to date on mechanistic interaction has been formulated assuming the potential outcome (PO) framework (Rubin, 1974) for causality or some essentially equivalent formulation, though the literature also offers some purely probabilistic approaches. The former category is exemplified by the stochastic PO approach of VanderWeele & Robins (2012); the latter is exemplified by previous work of the authors of this paper (Berzuini & Dawid, 2013) and by the recent paper of Ramsahai (2013). Section 8 discusses these approaches and their limitations. Aiming to overcome these limitations, we here supply a new definition of mechanistic interaction between the effects of possibly continuous causal factors \( A \in \mathcal{A} \) and \( B \in \mathcal{B} \) on a binary variable \( Y \). Our definition is in terms of departure from a null “generalised noisy-OR” model (Pearl, 1988; Lemmer & Gossink, 2004), under which the multiplicative causal effect of \( A \) (resp., \( B \)) on the probability of survival (\( Y = 0 \)) cannot be enhanced through manipulation of \( B \) (resp.,
Conditions are presented under which mechanistic interaction in the above sense can be assessed from observational data via simple tests on excess risk or superadditivity, under either a prospective or a retrospective sampling regime. The approach extends that of Berzuini & Dawid (2013) in not requiring any variable in the problem to depend on its causes in a deterministic way. In common with our previous work, the conditions for observational identifiability of mechanistic interaction are here expressed in terms of conditional independence relationships between problem variables, which can often be checked on a causal diagram representation of the problem. In §5 we establish a connection between mechanistic interaction and mediation by discussing sufficient conditions for a valid test of mechanistic interaction under a regime where a mediator of the effects of interest is fixed by intervention. In this case, what is being assessed is a mechanistic interaction between direct effects or, given a causal diagram representation of the problem, mechanistic interaction between path-specific effects. Finally, our approach caters for the many applied situations, e.g. in epidemiology, where A and B arise as continuous variables. The methods are illustrated by means of a study in experimental psychology.

2 Assumptions and notation

We are interested in the way a response variable Y would react to real or hypothetical manipulations of causal factors A and B, and in particular whether or not the effects of A and B on Y can be regarded as interacting in some fundamental mechanistic sense. In order to address this, we must first understand what might be meant by “no mechanistic interaction”. Here we suggest a possible explication of this concept. This however is not absolute, but relative to an appropriately chosen “context”. That is, we specify certain context variables W in the problem, which might modify in some way the dependence of Y on (A, B), and only consider this dependence within a fixed context, i.e., conditional on fixed values W = w.

In contrast to the formulation of Berzuini & Dawid (2013), we do not require that Y be a deterministic function of (A, B, W), but allow for a fully stochastic dependence of Y on these inputs. Note that this allows considerable freedom in the selection of the context variables. Indeed, even in those rare cases when there does exist a choice for W supporting a deterministic relationship, this might be regarded as inhabiting too deep a level of description to be useful for the purpose
at hand, and a more coarse-grained choice, yielding a genuinely stochastic relationship, could be more appropriate. In any given application, care must be taken to ensure that we are arguing at a suitable level of granularity. As an analogy, for most purposes it is appropriate to think of the determination of the sex of an embryo as governed by a random process (essentially a fair coin toss), even though a very detailed description of the positions, motions, properties and behaviours of the gametes prior to fertilisation might allow deterministic prediction.

Our definition of “no mechanistic interaction” will relate to a (possibly hypothetical) “interventional regime”, in which the values of $A$ and $B$ are set by some external agent or process. However, the data that available to investigate this property will generally have been generated under some other, typically purely observational, regime, where, in particular, the values of $A$ and $B$ have arisen in some uncontrolled stochastic way. We will need to be able to relate these regimes in order to transfer information from one to the other. To streamline this task we introduce the regime indicator $\sigma_{AB}$, a non-stochastic variable, where $\sigma_{AB} = ab$ indicates the interventional regime where $A$ is set to $a$ and $B$ to $b$, and $\sigma_{AB} = \emptyset$ the observational regime. More generally, $\sigma_X$ will denote a similar regime indicator for interventions on a set $X$ of variables of interest.

We introduce the symbol

$$\pi_w(a, b) := P(Y = 0 \mid W = w, \sigma_{AB} = ab), \quad (5)$$

for the probability of survival ($Y = 0$), in context $W = w$, when $A$ and $B$ are set to respective values $a$ and $b$ by an exogenous intervention. Then one way—which we shall henceforth adopt—of understanding the effect $(A, B)$ on $Y$ is by considering the dependence of $\pi_w(a, b)$ on $(a, b)$. In particular, the effect of changing the values set for $(A, B)$ from $(a', b')$ to $(a, b)$ is measured by the relative survival probability (RSP):

$$\frac{\pi_w(a, b)}{\pi_w(a', b')}.$$

2.1 Structural conditions

We shall impose the following structural conditions:

**Condition 1** The causal factors $A$ and $B$ are continuous or ordered categorical scalar random variables.
**Condition 2** The effects of $A$ and $B$ on $Y$ are positive: for any fixed $(b, w)$ (resp., $(a, w)$), $P(Y = 1 \mid W = w, \sigma_{AB} = ab)$ is a non-decreasing function of $a$ (resp., $b$).

An alternative expression of Condition 2 is that $\pi_w(a, b)$, given by (5), is, for each $w$, a non-increasing function of each of $a$ and $b$.

### 3 No mechanistic interaction

We henceforth make the structural assumptions of the previous section.

#### 3.1 Point null hypothesis

One possible way of expressing the concept of *no mechanistic interaction between $A$ and $B$ in producing $Y$* is that, for all $w$, we can express $\pi_w(a, b)$ in the product form

$$\pi_w(a, b) = \lambda_w(a)\mu_w(b)$$

for all $a \in A$, $b \in B$. We term this the *point null hypothesis*. Equation (6) can alternatively be expressed as requiring:

For all $a, a' \in A$, $b, b' \in B$,

$$\pi_w(a, b)\pi_w(a', b') = \pi_w(a, b')\pi_w(a', b)$$

(7)

or, when the denominators are positive,

$$\frac{\pi_w(a, b)}{\pi_w(a', b)} = \frac{\pi_w(a, b')}{\pi_w(a', b')}$$

(8)

or

$$\pi_w(a, b) = \frac{\pi_w(a', b')\pi_w(a, b')}{\pi_w(a', b')}.$$  

(9)

When both $A$ and $B$ are binary this is equivalent to the single requirement

$$\pi_w(1, 1) = \pi_w(1, 0)\pi_w(0, 1)/\pi_w(0, 0).$$

(10)

In the special case that $\pi_w(0, 0) = 1$, this becomes

$$\pi_w(1, 1) = \pi_w(1, 0)\pi_w(0, 1).$$

(11)
Imposing the further requirements $\pi_w(1, 0) = \pi_w(0, 1) = 0$ would now imply

$$\pi_w(1, 1) = 0,$$

and this constellation of values represents $Y$ as the Boolean expression $A \text{ OR } B$. The intermediate case of $\pi_w(0, 0) = 1$ together with (11) is the “noisy OR” generalisation of this, while the general expression (10) is “leaky noisy OR” (Pearl, 1988; Lemmer & Gossink, 2004; Zagorecki & Druzdzel, 2004).

Intuitively, the point null hypothesis asserts that manipulation of causal factor $A$ (resp., $B$) is an uncertain cause of $Y$ (it can produce the failure, but it will not always do so), and its effect on $Y$ will never depend on the value imposed on the remaining factor, $B$ (resp., $A$). For example, in a bowling club, imagine Player 1 being assigned a ball of size $A = a$ and Player 2 a ball of size $B = b$, and then each player being asked to try to knock his respective pin down. Let the event $Y = 1$ indicate that at least one of the two players knocks the pin over, and $Y = 0$ indicate instead that both pins “survive” the throw. Think of $\pi_w(a, b)$ as representing the probability of $Y = 0$ with assigned ball sizes $(a, b)$ with $W = w$ indicating specific circumstances such as air humidity, temperature. Then the point null hypothesis, as expressed by Equation (6), asserts that the probability of $Y = 0$ is the product of the probability $\lambda_w(a)$ that Player 1 fails to knock the pin down, and the probability $\mu_w(b)$ that Player 2 also fails. Which is to say that the performance of one player is not affected by the size of the ball given to the other player. Equation (8) makes the following implication explicit: giving one player a larger ball does not change the effect on the outcome one obtains by giving a larger ball to the other player.

### 3.2 Interval null hypothesis

Taking into account that we are only interested in synergistic (as opposed to antagonistic) interaction, we can weaken the above point null hypothesis, as expressed by (8) or (9), to yield the following interval null hypothesis:

For all $w$, and all $a \geq a' \in A$, $b \geq b' \in B$, we have

$$\frac{\pi_w(a, b)}{\pi_w(a', b')} \geq \frac{\pi_w(a, b')}{\pi_w(a', b')}.$$  \hspace{1cm} (13)

Equivalently,

$$\pi_w(a, b) \geq \frac{\pi_w(a', b) \pi_w(a, b')}{\pi_w(a', b')}.$$  \hspace{1cm} (14)
Informally: intervening to increase $B$ (resp., $A$) will not enhance the failure-inducing effect of an interventional increase in $A$ (resp., $B$).

3.3 Mechanistic interaction

**Definition 3.1** We say that the causal factors $A$ and $B$ exhibit mechanistic interaction, or that their effects interfere, in producing $Y$, when the interval null hypothesis (and so a fortiori the point null hypothesis) fails: that is to say, when there exists a value $w$ of the context variable $W$ and values $a > a'$ for $A$, and $b > b'$ for $B$, such that

$$\pi_w(a, b)\pi_w(a', b') < \pi_w(a', b')\pi_w(a, b').$$

(15)

When this holds, we may write $A*B$, or (to make explicit that the property is relative to the specified context variable $W$), $A*B[W]$. Note that, unlike statistical interaction, the property of mechanistic interaction is unaffected by monotonic increasing transformations of $A$ and $B$.

Because we focus on synergy, we have defined “interference to produce”; we could similarly define its antagonism counterpart, “interference to prevent”.

If Equation (15) holds then clearly $\pi_w(a', b) > 0$, $\pi_w(a, b') > 0$. Also $\pi_w(a', b') > 0$ for, if it were not so, all terms of the equation would be 0 by virtue of Condition [2]. Thus Definition 3.1 applies just when there exist $w$, $a > a'$, $b > b'$ such that

$$\pi_w(a, b)\pi_w(a', b') = \pi_w(a', b') \pi_w(a, b').$$

(16)

Inequality (15) represents a stochastic extension of the deterministic mechanistic interaction concept of Berzuini & Dawid (2013). Under such deterministic dependence of $Y$ on $(A, B, W)$, each term in (15) can only take values 0 or 1. Together with Condition [2] this implies:

$$\pi_w(a, b) = 0 < \pi_w(a', b) = \pi_w(a, b') = \pi_w(a', b') = 1.$$  

(17)

The above inequality says that there are values $b, b' \in B$ such that, in some context $W = w$, when we set $B = b$ a manipulation of $A$ from $a'$ to $a$ causes $Y$ to change from 0 to 1; whereas, in the same context, when we set $B = b'$, the same manipulation makes no difference to $Y$. In other words, whenever $Y$ is deterministic, presence of mechanistic interaction in our formulation is characterized by
the possibility of preventing the effect of a certain manipulation of $A$ by acting on $B$; and vice versa. If further $A$ and $B$ are binary, then $(a, a', b, b') = (1, 0, 1, 0)$, and the definition says that $A$ and $B$ interact mechanistically in producing $Y = 1$ when there exists a value $w$ of the context variable $W$ such that the dependence of $Y$ on $(A, B)$ obeys the Boolean conjunction law: $Y = A \land B$.

4 Observational identification of mechanistic interaction

We now consider how we might use observational data to assess the presence or absence of mechanistic interaction between the effects of $A$ and $B$ on $Y$. We shall do this be means of a set $C \subseteq W$ of observed context variables; the remaining variables $U = W \setminus C$ may be observed or unobserved.

We shall need to consider, in addition to the structural conditions of §2.1, some causal conditions, relating the behaviours under observational and interventional circumstances. These we express as follows, where we have used the symbol $\perp \perp$ for “conditionally independent of” (Dawid, 1979, 2002).

**Condition 3** $W \perp \perp (A, B, \sigma_{AB})$.

**Condition 4** $U \perp \perp (A, B, \sigma_{AB}) \mid C$.

**Condition 5** $Y \perp \perp \sigma_{AB} \mid (A, B, W)$.

Finally, we shall sometimes require observational independence between $A$ and $B$, conditional on $C$:

**Condition 6** $A \perp \perp B \mid (C, \sigma_{AB})$.

Condition 3 says that the context variables $W$ should be independent of the values of $(A, B)$, as well as of $\sigma_{AB}$, i.e., of whether those values arose naturally or by external intervention: this condition may be described as exogeneity. It is helpful for interpretive purposes, but is not required for the mathematical development; indeed, in §5 we shall find it fruitful to weaken it. Condition 4 says that, conditional on $C$, the distribution of $U$ is fixed: the same under both interventional and observational conditions and, further, independent of the values of $A$ and $B$. 

Condition 5 requires that the effects of \(A\) and \(B\) on \(Y\) be “unconfounded”, conditional on the context variables \(W\). Together with Condition 5, this defines \(W\) as a “sufficient covariate” for the effect of \((A, B)\) on \(Y\) (Guo & Dawid, 2010).

Condition 6 holds trivially for an interventional regime \(\sigma_{AB} = ab\), so only has force for the observational regime \(\sigma_{AB} = \emptyset\). It is a strong condition, but in certain circumstances can be avoided—see Corollary 1 below. Condition 6 is implied by Condition 3 and we shall want to retain this weaker condition.

4.1 Causal diagrams

It will sometimes be possible to use an influence diagram (ID) to represent the assumed causal and conditional independence relationships between the problem variables and the regimes (Dawid, 2002). This extension of the methodology of directed acyclic graphs (Cowell et al., 1999) incorporates the relevant regime indicators as decision nodes, sending arrows into the variables they refer to.

Figure 1, for example, might represent the effects of genetic variants \(A\) and \(B\) on myocardial infarction \(Y\) say, possibly mediated by obesity \(M\), with \(Z\) representing observed information about diet, social status, etc.. In the diagram, \(Z\) is assumed to be sufficient (albeit not minimal sufficient) for the effect of \(A\) and \(B\) on \(Y\). The diagram contains the regime indicator \(\sigma_{AB}\) (it also contains another regime indicator, \(\sigma_M\), which will be considered further in \(\S\) 6 below). Our causal conditions, expressing conditional independence relationships between the stochastic variables and the non-stochastic indicators, can be represented and checked using an ID, with the aid of a graphical criterion such as \(d\)-separation (Geiger et al., 1990) or its moralisation equivalent (Lauritzen et al., 1990). For a problem that can be modelled by the graph of Figure 1, causal conditions 3–6 follow by application of the moralisation criterion to the graph if we choose \(C\) to be empty and \(W \equiv (Z, U)\).

4.2 Main theorem

Our observational criterion for mechanistic interaction will involve a dichotomisation of the ranges of \(A\) and \(B\), determined by respective “cutoff thresholds” \(\tau_A\) and \(\tau_B\). Let \(\alpha\) be the indicator variable of “\(A > \tau_A\)”, and \(\beta\) the indicator variable of “\(B > \tau_B\)”. The symbol \(R_{ijc}\) is henceforth reinterpreted as:

\[
R_{ijc} = P(Y = 1 \mid \alpha = i, \beta = j, C = c, \sigma_{AB} = \emptyset),
\]
Figure 1: In an influence diagram such as this, variables may depend on their predecessors in the graph in a fully stochastic way. Regime nodes, here $\sigma_{AB}$ and $\sigma_M$, indicate whether the variables into which they send arrows are manipulated (interventional regime) or observed (observational regime).

and likewise $S_{ijc} = 1 - R_{ijc}$. We reinterpret the inequalities (1)–(4) correspondingly. Note that $R_{ijc}$ is estimable from data on variables $A$, $B$ and $C$, gathered under the observational regime.

**Theorem 4.1** Assume Conditions 1 and 2, and that Conditions 4–6 hold in stratum $C = c$. Suppose further that $\pi_{cu}(\tau_A, \tau_B) > 0$ for all $u$. Then the presence of superadditivity in stratum $C = c$ implies $A * B \ [W]$.

**Proof.** We proceed by assuming both superadditivity and the interval null hypothesis, and deriving a contradiction. In the following, all probabilities and expectations are taken under the observational regime $\sigma_{AB} = \emptyset$.

Using Condition 4, we have, for $i, j = 0, 1$:

$$S_{ijc} = E(S_{ijcu} \mid C = c)$$

where

$$S_{ijcu} := P(Y = 0 \mid \alpha = i, \beta = j, C = c, U = u).$$

Thus

$$S_{10c} + S_{01c} - S_{11c} - S_{00c} = E(S_{10cu} + S_{01cu} - S_{11cu} - S_{00cu} \mid C = c). \quad (19)$$
Let \( U^* := \{ u : S_{10}C + S_{01}C - S_{11}C - S_{00}C > 0 \} \). On account of (19) and (4), \( P(U^* \mid C = c) > 0 \); in particular, \( U^* \) is non-empty. Fix any \( u^* \in U^* \), and define \( S^*_{ij} := S_{ij}cu^* \), \( \pi^*(a|bc) := \pi_{cu^*}(a, b) \). Then
\[
S^*_{11} + S^*_{00} - S^*_{10} - S^*_{01} < 0. \tag{20}
\]

Now by virtue of Conditions 5 and 4,
\[
S^*_{ij} = E\{\pi^*(A, B) \mid \alpha = i, \beta = j, C = c\}. \tag{21}
\]

Let \( a > \tau_A \) and \( b > \tau_B \). By virtue of the interval null hypothesis (14) we then have
\[
\pi_0 \pi^*(a, b) \geq \pi^*(a, \tau_B)\pi^*(\tau_A, b) \tag{22}
\]
where \( \pi_0 := \pi^*(\tau_A, \tau_B) > 0 \) by assumption; hence
\[
\pi_0 S^*_{11} \geq E\{\pi^*(A, \tau_B)\pi^*(\tau_A, B) \mid \alpha = 1, \beta = 1, C = c\} \tag{23}
\]
\[
= E\{\pi^*(A, \tau_B) \mid \alpha = 1, C = c\} \times E\{\pi^*(\tau_A, B) \mid \beta = 1, C = c\}
\]
on using Condition 6. That is, defining
\[
S_A := E\{\pi^*(A, \tau_B) \mid \alpha = 1, C = c\}
\]
\[
S_{\overline{A}} := E\{\pi^*(A, \tau_B) \mid \alpha = 0, C = c\}
\]
\[
S_B := E\{\pi^*(\tau_A, B) \mid \beta = 1, C = c\}
\]
\[
S_{\overline{B}} := E\{\pi^*(\tau_A, B) \mid \beta = 0, C = c\},
\]
we have
\[
\pi_0 S^*_{11} \geq S_A S_B. \tag{24}
\]

Similarly,
\[
\pi_0 S^*_{00} \geq S_{\overline{A}} S_{\overline{B}}, \tag{25}
\]
while
\[
\pi_0 S^*_{01} \leq S_{\overline{A}} S_B, \tag{26}
\]
\[
\pi_0 S^*_{10} \leq S_A S_{\overline{B}}. \tag{27}
\]

From these inequalities and (20), we obtain \( (S_A - S_{\overline{A}})(S_B - S_{\overline{B}}) < 0 \). But by virtue of Condition 2 we have \( S_A \leq \pi_0 \leq S_{\overline{A}} \), \( S_B \leq \pi_0 \leq S_{\overline{B}} \), yielding \( (S_A - S_{\overline{A}})(S_B - S_{\overline{B}}) \geq 0 \). This contradiction proves the theorem. \( \square \)
Corollary 1 Suppose that both $A$ and $B$ are binary. Then the conclusion of Theorem 4.1 holds even if we remove the independence requirement of Condition 6.

Proof. In this case $\tau_A = \tau_B = 0$, and (23) becomes
\[
\pi_0 S_{11}^* \geq \pi^*(1,0)\pi^*(0,1)
= S_A S_B
\]
so that (24), and similarly (25), (26), (27), continue to hold even without assuming independence. The rest of the proof is unchanged. \qed

For the next Corollary we introduce the following weaker form of Condition 2 appropriate for cases where the directionality of the effect is not known \textit{a priori}:  

\textbf{Condition 7} The effect of $A$ on $Y$ is either positive, in the sense that $P(Y = 1 \mid W = w, \sigma_{AB} = ab)$ is a non-decreasing function of $a$ for all $(b, w)$; or negative, in the sense that $P(Y = 1 \mid W = w, \sigma_{AB} = ab)$ is a non-increasing function of $a$ for all $(b, w)$; and similarly with the rôles of $A$ and $B$ interchanged.

Corollary 2 Suppose that, in the statement of Theorem 4.1 or Corollary 1, we replace Condition 2 by the weaker Condition 7 and at the same time replace the superadditivity property (2) by the stronger excess risk property (1) (again reinterpreted in terms of definition (18)). Then the conclusion remains valid.

Proof. We use the same notation as in the proof of Theorem 4.1. Arguing similarly to that proof, we deduce that there exists a value $u^*$ of $U$ such that $R_{11}^* - R_{10}^* - R_{01}^* > 0$. This implies both
\[
S_{11}^* - S_{10}^* < 0
\]
and
\[
S_{11}^* - S_{01}^* < 0,
\]
as well as (19).

From (28), (24) and (27), we deduce $S_A(S_B - S_B^*) < 0$, whence $S_B < S_B^*$. But if the effect of $B$ were negative we would have $S_B \geq S_B^*$. Hence the effect of $B$ is
positive. Similarly, using (29), we deduce that the effect of $A$ is positive. The rest of the proof now follows as before. \hfill \square

**Comment.** By allowing the dichotomization of $A$ to be arbitrary, the above theorem fits the common situation where the continuous factor is made available in a dichotomized form, without the possibility of recovering the original continuous measurements.

## 5 Direct effects interaction

This section of the paper examines relationships between mechanistic interaction and mediation. Mediation analysis hinges on the concept of *direct effect* of a variable $X$ on $Y$. One variant of this concept, the *direct effect of $X$ on $Y$ controlling for $F$*, is meant to quantify the sensitivity of $Y$ to changes in $X$ when $F$ is held fixed by intervention, that is, when a hypothetical physical intervention changes the value of $X$ from some reference value $x$ to some value $x'$, while $F$ is set to some constant (Pearl, 2005; Robins & Greenland, 1992).

A connection between our account of mechanistic interaction and mediation can be established by defining the concept of mechanistic interaction between $A$ and $B$ when a further variable, $F$, is set by intervention to a constant. Let $Z := W \setminus F$ denote the unmanipulated context variables, and extend the notation (5) by writing

$$
\pi_f^z(a, b) := P(Y = 0 \mid Z = z, \sigma_{AB} = ab, \sigma_F = f), \quad (30)
$$

for the probability of $Y = 0$ given $Z = z$, conditional on $A$, $B$ and $F$ being manipulated to take on values, $a$, $b$ and $f$, respectively. Then take the the direct effect of $A$ on $Y$, in context $Z = z$, controlling for $(B = b, F = f)$, to be measured in terms of relative survival probability by the quantity

$$
\frac{\pi_f^z(a, b)}{\pi_f^z(a', b)}. \quad (31)
$$

Then a sensible way of defining “no direct (synergistic) mechanistic interaction” is to require that the act of setting $B$ to a higher value can never enhance the direct effect of $A$, as measured by (31) with $a > a'$. This leads to the following generalization of our previous Definition 3.1 of mechanistic interaction:
Definition 5.1  We say that the causal factors $A$ and $B$ interact mechanistically to produce $Y = 1$ under $F$-intervention if there exist values $(f, a > a', b > b', z)$ for $(F, A, B, Z)$, respectively, such that

$$\frac{\pi_f^z(a, b)}{\pi_f^z(a', b)} < \frac{\pi_f^z(a, b')}{\pi_f^z(a', b')}.$$  \hspace{1cm} (32)

In this case we write $A \ast B \mid F [Z]$ or, with the relevant context variable $Z$ understood, just $A \ast B \mid F$. When interest is attached to a specific value $f$ imposed on $F$, we may write $A \ast B \mid F = f [Z]$. We alternatively describe condition (32) as mechanistic interaction between the direct effects of $A$ and $B$ on $Y$, controlling for $F$, in context $Z = z$.

The following theorem holds.

Theorem 5.1  Suppose

Condition 8  $Y \perp \! \! \! \perp \sigma_F \mid (A, B, F, Z, \sigma_{AB})$.

(That is, conditional on $(A, B, F, Z)$, the dependence of $Y$ on $F$ is not further affected by the way the value of $F$ has been generated, be it by mere observation or by intervention.)

Then $A \ast B \mid F [Z]$ if and only if $A \ast B [F, Z]$.

Proof.  In this case $\pi_f^z(a, b) = \pi_{f_z}(a, b)$.  \hspace{1cm} \Box

Thus under Condition 8 we can use Theorem 4.1 and its Corollaries to investigate mechanistic interaction between direct effects. For such an application, the context variable $W$ must be replaced by $W^* = (W, F)$. Since typically $F$ will be affected by $(A, B)$, Condition 3 will no longer be appropriate after such replacement; but this will not affect the mathematical development, so long as we can still assume the weaker Condition 4.

6  Examples

We now illustrate our framework with the aid of four examples.
Example 1 Consider again the diagram of Figure 1. The diagram assumes

\[
M \not\perp \perp \sigma_{AB} \\
Y \not\perp \perp \sigma_M | \sigma_{AB}.
\]

Under the assumptions embodied in this graph, Condition 3–Condition 6 hold for \(W \equiv Z, C = \emptyset\). So if we can assume the validity of Condition 1 and Condition 2 (or Condition 7), the “total” mechanistic interaction \(A \ast B [U]\) can be examined under an observational regime by testing for the excess risk condition (1), or for the superadditivity condition (2), as appropriate, with \(C\) empty.

Also, it can be checked that Conditions 4–6 and Condition 8 hold for \(W \equiv (Z, M, U), C \equiv (Z, M)\) and \(F \equiv M\). So, under the relevant structural conditions, the excess risk and superadditivity conditions, in strata defined by \(M\) and \(Z\), provide a valid basis for testing for \(A \ast B | M [Z, U]\). Evidence in favour of the latter interaction might suggest that the two loci influence the disease through a mechanism that is independent of obesity.

\[\square\]
Example 2 In Figure 2(a) variable $U'$ is an unknown confounder of the causal effect of $M$ on $Y$. Yet, under the assumptions of this graph, Conditions 3–6 hold for $W \equiv (U', U'')$ and $C \equiv \emptyset$, and thus the causal conditions for an observational excess risk or superadditivity test of $A \ast B[U', U'']$ hold in this example. □

Example 3 our theory delegitimizes observational excess risk or superadditivity tests of the “direct effects” interaction $A \ast B \mid M$ in problems described by Figure 2(a). This is because the conditioning on $M$ creates a $d$-active path between the unobserved $U'$ and $(A, B)$, in violation of Condition 4.

A valid observational test of $A \ast B \mid M$ in problems with the above structure is possible if we are willing to make assumptions about the relevant distributions. We begin by noting that, according to Figure 2(a), the conditional distribution of the binary variable $Y$ given $(A, B, M)$ is independent of the regime under which the latter variables are generated. Then, for example, take $Y$ to obey

$$E(Y \mid A = a, B = b, M = m, U' = u, \sigma_{AB}, \sigma_M) = \exp(\beta_0 + \beta_A a + \beta_B b + \beta_M m + \beta_{AB} ab + \beta_{U'u}).$$  \hspace{1cm} (33)

Because $U'$ is unobserved, model (33) cannot be fitted to the data, nor its additive structure empirically tested. We thus integrate (33) with respect to $U'$. Under an observational regime the integration yields

$$R_{abm} = E(Y \mid A = a, B = b, M = m, \sigma_M = \emptyset)$$

$$= \exp(\beta_0 + \beta_A a + \beta_B b + \beta_M m + \beta_{AB} ab) \int \exp(\beta_{U'u}) p(u \mid a, b, m) \, du$$

$$= \exp\{g(a, b, m) + \beta_A a + \beta_B b + \beta_M m + \beta_{AB} ab\}. \hspace{1cm} (34)$$

This integration corresponds to deleting node $U'$ and the arrows it originates from Figure 2(a). The resulting “marginal” graph, shown in Figure 2(b), obeys the causal conditions for an observational test of $A \ast B \mid M$ in any stratum $C \equiv F \equiv M = c$ via excess risk or superadditivity, as appropriate. Now if we define

$$f(\beta) \equiv \exp(\beta_A + \beta_B + \beta_{AB}) - \exp(\beta_A) - \exp(\beta_B),$$  \hspace{1cm} (35)

then the excess risk (resp., superadditivity) condition for $A \ast B \mid M = 0$ under the causal assumptions of Figure 2(a) and the distributional assumptions of (33) can be expressed in the parametric form $f(\beta) > 0$ (resp., $f(\beta) + 1 > 0$).
Proceed by choosing appropriate distributions \( p(U') \) and \( p(M \mid A, B, U', \sigma_M = \emptyset) \) in such a way to complete the specification of an observational probability distribution over the graphs of Figure 2.

Considerations of identifiability of the distribution parameters will often suggest a discrete \( p(U') \), so that the overall distribution will have the form of a discrete mixture of distributions. Once this is done, Markov chain Monte Carlo methods can be used to generate samples from the posterior distribution of \( f(\beta) \), for a simulation-consistent estimate of the posterior probability of excess risk or superadditivity, as a basis for the test of \( A \ast B \mid M \). Analysis will generally involve trying different choices of \( M = 0 \) and a cautious adjustment for test multiplicity. We shall conclude in favour of a "direct effects" \( A \ast B \mid M \) interaction when the data adequately support presence of excess risk or superadditivity (as appropriate) for at least one choice of \( M = 0 \).

\[\Box\]

**Example 4** Brader and colleagues (Valentino et al., 2008) study the reaction of public opinion to media stories about immigration. A sample of 354 white, non-Latino, adults is invited to read a mock newspaper story about the costs of immigration. All the participants are given the same story, except that in some stories the immigrants are reported to be latino, and in the remaining ones they are white. Previous analyses of Brader’s data have been based on the causal assumptions of Figure 1 with \( A \) representing the participant’s age, the binary variable \( B \) indicating whether the participant was randomized to a “latino” or to a “white” story, \( M \) being an approximately normal variable representing the participant’s level of “anxiety about immigration”, as measured at the end of the reading through a questionnaire, and \( Z \) representing observed socio-educational variables independent of age. Let the binary variable \( Y \) indicate the participant’s response - yes or no - to the following question after the reading: “do you agree or disagree about the idea of sending Congress a letter of complaint about immigration policy?”. A positive response is interpreted in terms of opposition to immigration. Brader’s own analysis of the data (and a similar analysis by Imai et al. (2013)) suggests that white opposition to immigration is greater when the story involves latinos, compared to stories about white immigrants, and that a substantial component of this effect is mediated by anxiety.

Under the assumptions of Figure 1 a test of \( A \ast B[U] \) along the lines of Example 1 addresses the question whether the total effects of age \( A \) and ethnic connotation \( B \) interact mechanistically to produce an opposition reaction in the reader. A
positive answer to this question might be interpreted as saying that if we “allow the participant to become older”, before we test him, we shall cause his opinion to be less sensitive to ethnic connotation.

In order to probe deeper into mechanism, we may wish to test for $A \ast B \mid M [Z, U]$ along the lines of Example 1. This is a way of asking whether a young age and a latino connotation of the story would work in synergism to produce an opposition reaction if were to fix the level of anxiety to some constant level. A positive answer to this question might suggest that the reason why ethnicity has a lower impact on the opinions of the old is not because the old react with less anxiety, but because they let their rational (as opposed to emotional) thinking play a bigger role in the shaping of their opinion.

7  Causality and agency

The above example raises some issues of the interpretation of “causality” in our approach. According to our description so far, that concept has been closely tied to the possibility of making external interventions to set values for the “causal variables” $A$ and $B$. This conception is in line with philosophical “agency” theories of causality, Price (1991); Hausman (1998); Woodward (2003), which regard causes as handles for manipulating effects. However, such an anthropocentric manipulationist view is unnecessarily restrictive, and can hamper application of causal inference to numerous scientific disciplines that demand a more general notion of cause, not tied simply to what human agents can do.

In our application above, while the variable $A =$ “stated ethnicity” was manipulable (and was manipulated), we can not reasonably regard the variable $B =$ “age of participant” as manipulable. We might however conceive of being able to observe an individual at various points of her life, and be interested in the way in which her age then might make a difference to her psychological response to certain media framing techniques. Psychologists have knowledge, theories and hypotheses about the role of age in the response process. They can, for example, make informed guesses about - and explain on the basis of psychological theories - the different outcome we might have observed had the individual been younger or older than he is (e.g., “young people tend to react with less anxiety”). Specific psychological mechanisms and reactions are associated with young age. We should not give up looking into them simply because the age variable falls outside the standard manipulability theory of causation.
As another example, in epidemiology it is often appropriate to consider, as a cause of a disease, a variable such as genotype, whose manipulation by human beings is not practically possible; and application of mechanistic interaction tests to investigations of epistasis or pharmacogenomics will require a broader conception of “intervention” than the agency approach typically supplies. Recent discussions of the topic (Woodward, 2013) have loosened the strict confines of the manipulationist theory, regarding as an “intervention” any appropriate (in a sense that has to be made clear) exogenous causal process, without any necessary connection with human action.

8 Related work

A recent paper by VanderWeele & Robins (2012) (hereafter VR) tackles mechanistic interaction via stochastic (rather than deterministic) POs. In the standard PO formulation, the value that $Y$ would take in individual $\omega$ in response to an intervention that sets $(A, B)$ to values $(a, b)$ is regarded as a potential outcome, $Y_{ab}(\omega)$. Potential outcomes are fixed for each particular individual even before the treatment is applied, and unaffected by the particular regime in which the values of $A$ and $B$ are set. VR relax this by allowing each individual $\omega$ to be characterized by a stochastic potential outcome, $Y_{ab}(\omega)$, that varies in the individual according to a Bernoulli distribution with the expected value fixed by the intervention and by random circumstances, although these latter are assumed not to be affected by the treatment. Because of the latter constraint, it is not clear whether VR’s approach, as currently formulated, copes with situations where a stochastic mediator of the effect of $(A, B)$ on $Y$ introduces intervention-dependent random variation.

Ramsahai (2013) gives a fully probabilistic account of mechanistic interaction, which boasts aspects of greater generality relative to ours, including freedom from

---

1We also note that in Rubin’s standard PO formulation there is a value of the response for each individual and possible intervention, and such value is constant across all possible regimes, in the sense that it is not affected by the way the values of $A$ and $B$ are generated. In VR’s approach, the response has its expected value fixed by the particular individual, set of circumstances and intervention. But conditional on this expected value, is the realized value of the response assumed to vary across regimes? In other words, is the observationally detected response identical to what I would have observed had I intervened? And, if so, are the regime-specific versions of the response assumed independent? We feel that the question matters to the very purpose of carrying inferences from the observational to other regimes. These considerations are related to certain ambiguities of counterfactual-based formulations of causality (Dawid, 2000).
monotonicity assumptions about the effects of $A$ and $B$. There are also aspects of lesser generality: no attempt is made in Ramsahai’s paper to examine the implications of the presence of continuous causal factors. It is therefore appropriate to proceed by comparing Ramsahai’s method and ours in the special case where $A$ and $B$ are binary variables, with $(a, a', b, b') \equiv (1, 0, 1, 0)$. In this special case, our condition for presence of mechanistic interaction, as expressed by (16), specializes to

$$\frac{\pi_w(1, 1)}{\pi_w(0, 1)} < \frac{\pi_w(1, 0)}{\pi_w(0, 0)}.$$  \hfill (36)

As seen in §3.3, under Condition 2, Equation (36) implies strictly positive monotonicity of the effects of $A$ and $B$ upon $Y$, as expressed by the inequalities

$$\pi_w(1, 1) < \pi_w(1, 0),$$  \hfill (37)

$$\pi_w(1, 1) < \pi_w(0, 1).$$  \hfill (38)

In our approach, these inequalities are consistent with, but not sufficient for, the presence of mechanistic interaction. In fact, consistently with our concluding remarks of §3.1, inequalities (37)–(38) do not imply (36). By contrast, in Ramsahai’s approach, those inequalities are taken to define mechanistic interaction for binary variables. Hence Ramsahai’s definition of mechanistic interaction is weaker than ours. The more exacting nature of our definition of mechanistic interaction, combined with allowance for continuity, explains the stronger assumptions required in our approach compared to those of Ramsahai.

To elucidate the differences between the approaches, suppose that, in the bowling example, $A$ (the first player’s ball size) takes value 0 (the player has no ball to throw) or 1 (the player throws a ball). Interpret $B$ analogously. It then seems reasonable, on the basis of physics and common sense, to assume that inequalities (37)–(38) hold in this example. In Ramsahai’s formulation, this is sufficient to conclude in favour of mechanistic interaction between the effects of the throws of the two players, even before looking into the data, and even if the two players act independently. This appears to clash with our psychological notion of synergism. By contrast, in our formulation, (37)–(38) are not sufficient to conclude in favour of mechanistic interaction: we would need to find evidence that contradicts the idea of independent throws expressed by (14)—which accords with our intuition.

To conclude, we note that our approach uses statistics (excess risk and superadditivity) which are often testable at negligible computational cost in prospective studies, and (approximately and under assumptions) also in retrospective studies.
By contrast,, attention needs to be paid to the computational feasibility of Ramsahai’s approach.

9 Discussion

Mechanistic interaction has often been tackled within a potential outcome framework (Rubin, 1974) or within an equivalent formulation of causality. We have discussed possible limitations of this approach. We have also discussed limitations of current approaches to mechanistic interaction which reject the potential outcome formulation in favour of the standard probability formalism. Motivated by the limitations of the previous approaches, we have proposed a novel definition of the causal notion of mechanistic interaction, and presented sufficient conditions for its identification from observational data. Because these conditions are expressed in terms of conditional independence, they hold irrespective of particular parametric or distributional assumptions about the problem variables. A further advantage of our conditional independence formulation of the identifiability conditions is that these can be straightforwardly checked on a causal diagram of the problem, when this is available. The use of causal diagrams for the mentioned purposes has been extensively illustrated.

We recommend the use of our conditions as a first step in an analysis of mechanistic interaction. In applications where our conditions do not directly support the desired test, one can attempt to make the mechanistic interaction of interest observationally identifiable by introducing appropriate distributional assumptions. We have discussed and illustrated this in Example 3.

Our theory provides conditions for testing for mechanistic interaction in (real or hypothetical) situations in which an intervention is exerted on variables (even post-treatment ones) different from the main factors A and B of interest. We have discussed the connection between this and the idea of mechanistic interaction between effects that flow along specific paths in a causal diagram representation of the problem.

Importantly, our method does not require the assumption that Y depends on its causal influences in a functional way. By relaxing such an assumption, our method gains applicability in a much wider range of situations, and confers more leeway on the researcher in the choice of the conditioning variables in the test.
Once the conditions for a test of the mechanistic interaction of interest have been found valid, the actual test involves simple (and well known) excess risk or superadditivity statistics. These tests are valid under prospective sampling and (under assumptions) retrospective sampling. In the latter case, a key assumption is that the response event of interest is rare under any possible configuration of the causal factors. In the context of retrospective case-control studies in epidemiology, this is the well-known rare disease assumption that typically motivates this kind of studies.

Finally, our approach embraces the very large class of applications where the main causal factors, $A$ and $B$, are only available as a discretized version of the fundamental variables, no longer available in their original continuous form.

Various possible enhancements of the method are envisaged, one of these being the extension of the theory to embrace higher-order mechanistic interactions. Equally important will be the application of the method in a variety of situations and disciplines, from genetic epidemiology (e.g., in the identification of gene-environment interactions) to experimental psychology. We hope that the proposed method will help researchers better to identify from data analysis small sets of interactions underlying mechanisms of scientific interest.

10 Acknowledgments

Carlo Berzuini was partially supported by the FP7-305280 MIMOmics European Collaborative Project, as part of the HEALTH-2012- INNOVATION scheme.

References

BERNARDINELLI, L., BERZUINI, C., FOCO, L. & PASTORINO, R. (2012). Ion Channels as a Possible Mechanism of Neurodegeneration in Multiple Sclerosis. John Wiley and Sons, Ltd, pp. 208–217.

BERZUINI, C. & DAWID, A. P. (2013). Deep determinism and the assessment of mechanistic interaction. Biostatistics 14, 502–513.

COWELL, R. G., DAWID, A. P., LAURITZEN, S. L. & SPIEGELHALTER, D. J. (1999). Probabilistic Networks and Expert Systems. Springer.
DAWID, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* 41, 1–31.

DAWID, A. P. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association* 95, 407–424.

DAWID, A. P. (2002). Influence diagrams for causal modelling and inference. *International Statistical Review* 70, 161–189. Corrigenda, *ibid.*, 437.

GEIGER, D., VERMA, T. & PEARL, J. (1990). Identifying independence in Bayesian networks. *Networks* 20, 507–534.

GREENLAND, S. & POOLE, C. (1988). Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work, Environment and Health* 14, 125–129.

GUO, H. & DAWID, A. P. (2010). Sufficient covariates and linear propensity analysis. *Journal of Machine Learning Research Workshop and Conference Proceedings* 9, 281–288. Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna, Sardinia, Italy, May 13–15, 2010, edited by Y. W. Teh and D. M. Titterington. http://jmlr.csail.mit.edu/proceedings/papers/v9/guo10a/guo10a.pdf.

HAUSMAN, D. (1998). *Causal Asymmetries*. Cambridge: Cambridge University Press.

IMAI, K., TINGLEY, D. & YAMAMOTO, T. (2013). Experimental designs for identifying causal mechanisms (with Discussion). *Journal of the Royal Statistical Society, Series A* 176, 5–51.

LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. & LEIMER, H. G. (1990). Independence properties of directed Markov fields. *Networks* 20, 491–505.

LEMMER, J. F. & GOSSINK, D. E. (2004). Recursive noisy or - a rule for estimating complex probabilistic interactions. *Trans. Sys. Man Cyber. Part B* 34, 2252–2261.

PEARL, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
PEARL, J. (2005). Direct and indirect effects. In Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence. San Francisco, CA: Morgan Kaufmann.

PRICE, H. (1991). Agency and probabilistic causality. British Journal for the Philosophy of Science 42, 157–176.

RAMSAHAI, R. R. (2013). Probabilistic causality and detecting collections of interdependence patterns. Journal of the Royal Statistical Society, Series B 75, 705–723.

ROBINS, J. & GREENLAND, S. (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology 3, 143–155.

ROTHMAN, K. J. (1976). Causes. American Journal of Epidemiology 104, 587–592.

ROTHMAN, K. J. & GREENLAND, S. (1998). Modern Epidemiology. Philadelphia: Lippincott-Raven.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66, 688–701.

VALENTINO, N. A., SUHAY, E. & BRADER, T. (2008). What triggers public opposition to immigration?: Anxiety, group cues, and immigration threat. American Journal of Political Science 52, 959–978.

VANDERWEELE, T. J. (2009). Sufficient cause interactions and statistical interactions. Epidemiology 20, 6–13.

VANDERWEELE, T. J. (2010). Epistatic interactions. Statistical Applications in Genetics and Molecular Biology 1, 1–22.

VANDERWEELE, T. J. (2011). Empirical tests for compositional epistasis. Nature Reviews Genetics 11, 166.

VANDERWEELE, T. J. & LAIRD, N. M. (2011). Tests for compositional epistasis under single interaction-parameter models. Annals of Human Genetics, Special Issue on Epistasis 75, 146–156.

VANDERWEELE, T. J. & ROBINS, J. M. (2008). Empirical and counterfactual conditions for sufficient cause interactions. Biometrika 95, 49–61.
VanderWeele, T. J. & Robins, J. M. (2009). Minimal sufficient causation and directed acyclic graphs. *Annals of Statistics* 37, 1437–1465.

VanderWeele, T. J. & Robins, J. M. (2012). Stochastic counterfactuals and sufficient causes. *Statistica Sinica* 22, 379–392.

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

Woodward, J. (2013). *Causation and Manipulability*. Stanford.

Zagorecki, A. & Druzdzel, M. (2004). An empirical study of probability elicitation under noisy-or assumption. In *Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, V. Barr & Z. Markov, eds. AAAI Press.