New Advances in Causation, Agency and Moral Responsibility

New Advances in Causation, Agency and Moral Responsibility

Edited by

Fabio Bacchini, Stefano Caputo and Massimo Dell'Utri

CAMBRIDGE SCHOLARS

PUBLISHING

New Advances in Causation, Agency and Moral Responsibility, Edited by Fabio Bacchini, Stefano Caputo and Massimo Dell'Utri

This book first published 2014

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data A catalogue record for this book is available from the British Library

Copyright © 2014 by Fabio Bacchini, Stefano Caputo, Massimo Dell'Utri and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-6625-3, ISBN (13): 978-1-4438-6625-5

TABLE OF CONTENTS

Introduction vii
Chapter One 1 Causation and the Interventionist Vector of Explanatory Depth Fernanda Samaniego
Chapter Two
Chapter Three
Chapter Four
Chapter Five
Chapter Six
Chapter Seven
Chapter Eight

Table of Contents

Chapter Nine Responsibility Regardless of Causation Federico L.G. Faroldi	121
Chapter Ten Skepticism and Control Sofia Bonicalzi	144
Chapter Eleven Understanding Strength of Will Michael Brent	165
Chapter Twelve Responsibility and Narrative Agency Irene Bucelli	179
Chapter Thirteen Empathy and Concern for Others Sarah Songhorian	194
Chapter Fourteen Truth and Moral Responsibility P. Roger Turner	214
Chapter Fifteen "I was not aware I was doing that" Franziska Müller	231
Abstracts	246
Contributors	253
Index of Names	256

INTRODUCTION

This volume brings together a number of previously unpublished essays that aim to advance our philosophical understanding of specific aspects of causality, agency and moral responsibility. The first group of contributions focuses on causality. One of the most recent and influential proposals in accounting for causal explanations is the interventionist theory developed by James Woodward, and the first two essays deal with the notion of 'explanatory depth' developed by Woodward and Hitchcock within interventionism.

Fernanda Samaniego argues that interventionism may lead to counterintuitive results when it is used in an attempt to define a notion of explanatory depth. In particular, if we apply the interventionist criteria of explanatory depth to cases with multiple correlated causes, such as the case of the causal relationship between cholesterol and atherosclerosis (also known as arteriosclerotic vascular disease), simpler explanations may turn out to be *deeper* than more detailed ones. Samaniego claims that we do have good reasons to rescue interventionism from this difficulty, and provides a new interventionist definition of explanatory depth that avoids it.

Alexandre Marcellesi goes further in maintaining that explanatory depth, as developed by Woodward and Hitchcock, is not an adequate account of the factors that make one causal explanation better than another. He claims that explanatory depth, for one thing, is inconsistent with the plausible and perhaps inescapable view that causal explanations are better when they refer to causes that are proportional to their effectsthat is, causes that are both necessary and sufficient for their effects. He goes on to argue that explanatory depth is not even an explanatory notion; in fact, after reminding us that explanatory depth should be one explanatory virtue of causal explanations, to be distinguished from ordinary theoretical virtues, such as predictive power, he shows that it is actually no more than predictive power in disguise. Marcellesi assumes that causal explanations and causal predictions are two at-leastconceptually distinct things-an assumption on which philosophers of science are perhaps required to focus their attention anew by the interventionist theory and its troubles.

In the context of the debate about the use of racial categories in

biomedicine, Ludovica Lorusso can be read as stressing the same point: since causal predictions are not the same thing as causal explanations, we should not confuse the predictive use—both causal and noncausal—of self-identified races in biomedical research with their putative causal explanatory role. Lorusso argues against the presumption that the only way of causally explaining different risks of developing a complex disease between different self-identified racial groups consists in postulating causally relevant genetic differences between the groups. In fact, there is no biological evidence that such causally relevant genetic differences exist. Interestingly, self-identified races are claimed to be not only possibly correlated to non-genetic causes of an increased risk of developing a complex disease—which would make races useful proxies in biomedicine even if there are no genetic causes of an increased risk that are correlated to them—but even possible causes of those non-genetic causes—what would make races non-genetically causally relevant to the increased risk.

There are two papers which deal with the metaphysics of causal relations rather than with the epistemology of causal explanations. Lorenzo Azzano examines what it is for something to be an interference in a causal process. After determining that an interference is required not to directly alter either the cause or the effect of the causal process it alters, he concludes that interferences can only alter causal processes that admit intermediate steps. Thus showing that all causation is sufficient causation is actually "tremendous work" consisting of sustaining some very robust metaphysical claims about either simultaneity, transitivity or the micro/macro relationship. For instance, one way of defending causal sufficiency is holding that all breakable causation just is macrocausation that can be exhaustively reduced to two-event discrete unbreakable microcausation.

Another interesting notion concerning causal interaction besides interference is that of coincidence. In her paper, Alessandra Melas deals with absolute coincidences and shows how a common cause model—i.e. Salmon's interactive fork model—can account for them. Seeing absolute coincidences as *x*-shaped interactive forks—in Salmon's terms—reconciles this particular aspect of chance with the Principle of Causality.

An interesting question that has become popular in the philosophy of perception is that of whether we can perceive some higher-order relation as causation. Various positions have emerged concerning our alleged capacity to *visually* perceive causal relations, but little space has been devoted to answering the same question for other sense modalities. Elvira Di Bona takes into account a few versions of the phenomenal contrast method developed by Susanna Siegel in order to isolate one version entitling us to affirm that, after all, we can *auditorily* perceive a causal relation. The main difficulty, here, is putting together the epistemological and the phenomenological perspectives.

Causation, agency and moral responsibility are of course deeply intertwined notions, and a large proportion of the volume is taken up by papers that attempt to shed light on their mutual connections or to defend certain claims concerning them. For instance, according to Carolina Sartorio, moral responsibility results from a certain kind of the agent's freedom with regards to X which in turn supervenes on the actual causal sequence producing X. Thus moral responsibility is solidly grounded on causation: there is no responsibility difference without a causal difference. Sartorio shows that some apparent counterexamples to the supervenience thesis actually end up by supporting it. Moreover, in developing her defensive argument Sartorio explicitly takes causation to be a vehicle for the transmission of an agent's moral responsibility from one outcome to another. A causal link between two outcomes is by itself not sufficient to transmit responsibility, as for instance some relevant epistemic conditions in the agent must obtain; a causal relation, however, turns out to be a necessary condition.

While many philosophers have deemed it natural to explain responsibility in causal terms-however different the details of the explanation might be—Santoro and Di Paola seem to reverse the direction of the explanation when they propose a Wittgensteinian view of causal ascriptions based on the more primitive language game of blame ascriptions. In what they call the scapegoat theory of causality, in fact, causes are pointed at just as scapegoats for explaining otherwise mysterious events; in this view, what we are doing when we point at a cause is just *blaming* some event for producing another event. According to their pragmatic analysis (as developed by Robert Brandom), the deployment of causal vocabulary is only possible if an inferential discursive practice of blame and responsibility ascription is presentwhich in turn requires the possession of an intentional vocabulary of blame and responsibility. To support their view, Santoro and Di Paola argue that we can express causal relations purely by means of intentional vocabulary, but not the reverse: we can "blame a cause", while we cannot "cause a blame".

Taking a different path, Federico Faroldi argues that responsibility is independent from causation—and he cites group responsibility, shared responsibility and vicarious responsibility as examples of responsibility that are obtained from requisites other than causal. Although he refers to legal rather than moral responsibility, his thesis can easily be reformulated as opposing the widely accepted idea that moral responsibility requires causing as a necessary condition at least.

In her essay, Sofia Bonicalzi attempts to map the different characterizations that compatibilists and libertarianists have offered of the nature of the conditions that must be guaranteed in order for moral responsibility to be obtained. While compatibilists usually admit a causal relation as a necessary condition for moral responsibility, libertarians have often tried to say that a necessary condition for being responsible-that is, being *free*—requires a departure from causal relations regulating the physical world, at least as long as causal relations are meant to be deterministic. Bonicalzi says that a decisive notion in both approaches is that of *control*, and examines how it is treated within theories positioned at either side of the boundary. Among libertarians, for example, Robert Kane has proposed to ground the agent's control-via *ultimate responsibility*on causal chains originating from a certain number of indeterministic, causally sufficient self-forming actions. Bonicalzi finally advocates a sceptical solution to free will in the footsteps of Derk Pereboom that in her opinion can rescue deontic morality through dropping the ideas of desert and blame and adopting a revisionist forward-looking conception of moral judgments.

While this group of essays in the volume focuses on the relations between causation and responsibility, some others deal with the connections between causation and agency, on the one hand, and those between agency and responsibility, on the other. Consider Michael Brent's contribution. Here the problem is to explain what happens when an agent succeeds in persisting with a resolution in the face of a compelling desire to the contrary. According to Richard Holton, whose arguments are examined by Brent, neither a traditional Humean account (where all action is explained in terms of beliefs and desires) nor a modified Humean account (where intentions, besides desires and beliefs, are considered in the explanation of actions) can adequately clear up this quandary. Holton's solution-which, Brent argues, is not immune to serious difficultiesconsists of postulating another irreducible motivational factor, namely, that of *willpower*. What is worth noting is that willpower can be seen as the capacity to resist motivational interference due to powerful desires threatening to undermine previous resolutions. The analogy between causal interference and unbreakable processes (Azzano), on one side, and motivational interferences and "unbreakable" resolutions (Brent), on the other, is tempting. Moreover, when Brent turns to the difficulties affecting Holton's view, he emphasizes that Holton offers no explanation of what the *causal role* of willpower is and by what causal process it can make the

agent capable of refraining from acting according to the desire threatening her previous resolution. Holton does say that willpower consists of recalling the resolution and refusing to revise it; we have no idea, however, about how these mental acts should make us capable of resisting temptations (e.g., whether by strengthening the resolution or by weakening or blocking the desire to the contrary, or something else entirely), nor are we given any help to clarify whether these mental acts are causally necessary or not to displaying strength of will.

Irene Bucelli is interested in evaluating what consequences reflective endorsement theories of agency have on responsibility. She considers Velleman's narrative model of agency particularly promising for providing maybe necessary and sufficient-conditions necessarv-and for responsibility, since "what the narrative model ends up identifying as action already has the two standard conditions for responsibility readily built in"-the two conditions being a causal ("volitional") and an epistemic one. Yet Bucelli concludes that the project fails inasmuch as omissions can reveal themselves to be acts of negligence we are responsible for. In fact, it appears that no reflective endorsement theory can easily accommodate omissions as instances of full-blown agency. In short, as the range of actions that the model is capable of accounting for just includes deliberated actions, it necessarily leaves out non-deliberated omissions we are routinely considered responsible for.

Another interesting project is that of Sarah Songhorian. Her aim is to single out a minimal concept of empathy common to all different forms of empathic capacities, and determine its causal relevance to behaviour and, in particular, to moral behaviour. She concludes that while this minimal concept of empathy has neither the necessary force to produce moral behaviour nor to entail consequences about responsibility and other normative categories, the concept of *sympathy* can start doing the job thanks to its metaethical dimension, and this in particular may provide a narrow path from the *is* of empathy to the *ought* of normative attitudes.

According to P. Roger Turner, some profound consequences for discussions on moral responsibility arise from certain facts about the nature of truth. In particular, the obvious truism that truth depends on the world—i.e. that, for every true proposition p, 'p' is true *because* it is the case that p—entails that counterexamples to one of the main incompatibilist arguments are impossible, and that compatibilism fails. One of the remarkable corollaries of Turner's defence of incompatibilism is the thesis that, if an agent is directly morally responsible for what p's truth depends on (in the sense of 'depends on' in which truth depends on the world), then the agent is directly morally responsible for p's truth at

Introduction

least partially. This thesis provides us with a guide to what we may call the compositionality of responsibility: Turner argues that if the agent is morally responsible for a single state of affairs that a complex proposition's truth depends upon—in the sense, for example, that the agent is morally responsible that s, and the truth of the complex proposition 't entails s' compositionally depends upon s—then the agent is also morally responsible for the truth of the complex proposition (t entails s') and is even morally responsible *that* the corresponding relation between states of affairs holds (e.g., *that* t entails s).

While Turner's approach has the important characteristic of weakening in a very particular way the connection between moral responsibility and the epistemic conditions usually required for moral responsibility to be obtained—because for him an agent who is morally responsible that s also in virtue of her knowing both that s and that she is causing s, can be *ipso facto* morally responsible that v although she knows neither that v nor that she is causing v, provided that v's truth depends on s-Franziska Müller wants to ascertain to what extent another connection that many philosophers have considered to be a very strong one-i.e. that between awareness of one's doing and intentional action—stands. Thus she puts to the test Elizabeth Anscombe's famous statement that if someone is not aware she is doing X, doing X is not an intentional action of hers. First, Müller establishes that there are many different ways in which we may interpret Anscombe's claim depending on what exactly the subject is supposed to be unaware of. Then-after picking out the most plausible option-she examines the numerous senses in which the subject can be said to be (and, not to be) aware of what she is doing. Her main thesis, here, is that having a dispositional belief to be doing X is not sufficient to grant awareness of one's doing X.

Philosophers still have a great deal of work to do before such complex notions as causality, agency and moral responsibility will be fully understood. We believe, however, that these essays constitute a valid contribution towards our comprehension of such matters.ⁱ

Sassari (Italy), March 2, 2014

Fabio Bacchini Stefano Caputo Massimo Dell'Utri

Notes

ⁱ We would like to thank Silvia Negroni for her invaluable assistance and helpful advices.

CHAPTER ONE

CAUSATION AND THE INTERVENTIONIST VECTOR OF EXPLANATORY DEPTH

FERNANDA SAMANIEGO

1. Introduction

One of the most recent and influencing proposals in modelling causal scientific explanations is the Interventionist Theory or *interventionism* (James Woodward 2003). This theory aims to be applicable to causal explanations from a broad variety of natural, social and health sciences.

Even though the wide applicability of Woodward's interventionism has become evident (see for example Waters 2007 or Suárez & San Pedro 2011), there is still some skepticism towards the suitability of some interventionist notions in specific cases (see Marcellesi 2010; Phyllis *et al* 2011; Russo 2012; Samaniego 2013; Reutlinger 2013; Baumgartner & Glynn 2013). This skepticism is, among other reasons, due to the lack of applicability of the interventionist notion of *explanatory depth* to cases where several causes are intertwined.

Given the difficulties in applying interventionism to some relevant cases, one possible reaction would be to criticize and dismiss the theory. The alternative reaction, here adopted, is attempting to improve interventionism in such a way that it can account for the complex causal explanations postulated by scientists.

The main objective of this paper is to improve interventionism by proposing a new notion of explanatory depth. It will be argued that this new notion is suitable to assess multi-causal patterns difficult to evaluate with Woodward's original notion of explanatory depth. The content of the paper will be organized as follows: the central notions of interventionism will be presented in section 2, and the three criteria of explanatory depth will receive special attention. In the next section these three criteria will be applied to a particular case: the causal relationship between cholesterol

Chapter One

and atherosclerosis. In section 4 it will be discussed how the interventionist criteria of explanatory depth, as originally proposed in the interventionist theory, are not equally fulfilled in the cholesterol-atherosclerosis case of study. This will lead us to propose (in sec.5) a new way of understanding the interventionist notion of explanatory depth as a three dimensional vector. Section 6 presents some final remarks.

2. The Interventionist Theory

The essence of a causal explanation, according to the interventionist theory, consists in exhibiting a pattern of counterfactual dependence associated with relationships that are potentially exploitable for purposes of manipulation and control (see Woodward, 2003, pp. 13-16). In other words, a basic idea of interventionism is that causal scientific explanations do not aim simply at satisfying our intellectual curiosity, but are often guided by the goal of finding information potentially relevant for the *manipulation* and *control* of the explained events.

2.1. The notion of cause

The very notion of cause in the interventionist theory is linked to manipulation. It is defined as follows:

 \hat{C} is a genuine cause of the effect E if, given the appropriate background conditions, there is a possible manipulation of the cause C so that this is also a way of manipulating or changing the effect E (see Woodward 2003, sec.2.2, or Woodward 2008, sec.1).

In other words, causal relations entail some changes upon the values of E whenever the values of C are modified. The changes of value performed over C must be reproducible in the sense that responses to the effect E must be in some way repetitive or systematic.

The fact that a set of counterfactuals is associated with every causal relationship is also essential in the interventionist theory. When C is a cause of E, the associated counterfactuals will be of the following kind: "If C were manipulated by the intervention I, then E would experiment such and such changes".

2.2. Formal definition of intervention

The notion of intervention is formally defined in the interventionist theory as follows

(IN) I assuming some value I = i is an intervention on C with respect to

E if and only if I = i is an actual cause of the value taken by *C*, and *I* meets the following conditions:

(IN-i) I must be the only cause of C; i.e., the intervention must completely disrupt the causal relationship between C and its previous causes so that the value of C is set entirely by I.

(IN-ii) I should not itself be caused by any cause that affects E via a route that does not go through C.

(IN-iii) I must not directly cause E via a route that does not go through C.

(IN-iv) *I* leaves the values taken by any causes of *E* except those that are on the directed path from *I* to *C* to *E* (should this exist) unchanged (Woodward, 2010, sec. 5; and 2003, p. 98).¹

2.3. The notion of invariance

The notions of causation, explanation and invariance are closely intertwined in the interventionist theory. According to this theory, a generalization G—i.e. a relationship expressing a putative causal connection between two variables—counts as causal or explanatory if and only if it is invariant under some appropriate set interventions (see Woodward 2003, p. 15; p. 239). Like the notion of intervention, the notion of invariance is modal in the sense that it tells us whether a putative causal relationship would remain stable if, perhaps contrary to fact, certain changes or interventions were to occur.

A generalization G relating changes in the cause C (from the value c to the value c') to changes in the effect E (from e to e') is invariant under a testing intervention I if and only if G correctly describes what the new value of E, e', would be under this change; that is, if and only if it remains true that G(c') = e' (see Woodward 2003, sec. 6.2).

2.4. The notion of explanatory depth

According to the interventionist theory, the import of a given explanation relies on its capacity to provide answers to counterfactual questions, i.e., questions about what would happen under circumstances different to the actual ones (from now on *what-if-questions*). And in satisfactory causal explanations, the patterns of dependence between causes and effects (or generalizations) are invariant under a set of interventions.

The notion of explanatory depth is introduced in the interventionist theory using the following example (formulated by Haavelmo 1944, pp.

27-28). Suppose that two different explanations are offered about how to increase and decrease the speed of an automobile. The simplest explanation relates the speed to the distance of the gas pedal from the bottom of the car. And a second and more elaborated explanation details the whole inner mechanism of the car, tells us how the motor and the carburettor work and so on.

Both explanations are successful in terms of providing sufficient information to operate the vehicle. Besides, both explanations are valid under the criteria of the interventionist theory because they identify patterns of counterfactual dependence that enable to control the speed of the automobile. In the first case, the postulated relation between speed and the gas pedal remains invariant under some interventions, for example, under changes in the values of the pedal inclination. However, this relation postulated by the simplest explanation fails if the car runs out of petrol, or if any element inside the car does not work properly. The second explanation, in contrast, is invariant under interventions on any part of the mechanism. And thus it answers a larger number of *what-if-questions*. Among other things, it explains why the car does not accelerate if the gas tank is empty. Therefore, in interventionist terms, the second explanation is deeper than the first one.

The notion of explanatory depth is not properly defined in the interventionist theory in the sense that the theory does not provide us with a set of necessary and sufficient conditions for this notion. Nevertheless one may identify three criteria that Woodward's interventionism takes into account in order to assess explanatory depth. I turn now to specify those three criteria.

Criterion I. A causal explanation is explanatorily deep if the generalization figuring in the explanation is invariant under a wide range of interventions (see Woodward 2003, p. 311).

Suppose we define the variables that figure in the explanation of the car's acceleration as P=pressing the pedal, and A=acceleration. If the pedal is pressed then P=1, and if the pedal is not pressed P=0. Similarly, A=1 if the car accelerates and A=0 if the velocity of the car is constant. And suppose we have an alternative explanation that also appeals to the causal relation between P and A but, instead of defining P and A as bi-valued variables, it defines them as multi-valued variables. P can take any real value in the possible range of the pedal's inclination, and A can take any value between $1m/s^2$ and $1000m/s^2$. Criterion I would then tell us that the explanation appealing to bi-valued variables, because it shows "how any

one of a great number of changes in the *explanans* variables will lead to one of many possible changes in their *explanandum* variables. In other words, [it] gives us information about a much more detailed and fine-grained quantitative pattern of counterfactual dependence than the 'binary' pattern" (Woodward 2003, p. 206). This means that there are many more ways to intervene upon the multi-valued variable P than upon the bi-valuated P.

Criterion II. A causal explanation is explanatorily deep if the generalization figuring in the explanation is invariant, not only under a wide range of interventions (criterion I), but also under a wide variety of different kinds of interventions (see Woodward 2003, pp. 211-215).

Criterion II may be illustrated again with the example of the car. In fact, Woodward (2003, pp. 259-260) uses this example to illustrate that the second and elaborated explanation is deeper than the simple one because the generalizations figuring in the elaborated explanation are invariant under a wider variety of different interventions. In addition to being able to intervene upon the pedal in many ways, the elaborated explanation allows us to perform many new interventions (upon the motor, the carburettor, the petrol tank, etc.) that were not considered in the first and simple explanation. Therefore, according to the interventionist theory, it is a deeper causal explanation of the car's acceleration.

This invariance under a variety of kinds of interventions is reflected, in turn, in the diverse and detailed counterfactuals associated to that explanation. According to the interventionist theory in order "to elucidate certain kinds of causal claims, including claims about direct causal relationships and singular causal claims, one must appeal to counterfactuals with detailed antecedents—counterfactuals that describe what will happen under combinations of manipulations or interventions, rather than under single manipulations" (see Woodward 2003, p. 21).

Criterion III. A causal explanation is explanatorily deep if it is able to answer a wide range of counterfactual questions about the conditions under which the *explanandum* would have been different (see Woodward, 2003, p. 191). In other words, the deeper the explanation, the wider the range of "what-if-things-had-been-different" questions it answers (see Woodward 2003, p. 311).²

Again our original example of the car illustrates how the second explanation is deeper than the fist one. The second explanation is able to answer what would happen if the carburettor breaks, if the petrol does not flow from the tank to the motor and so on. Whereas the first explanation is able to answer one single *what-if-question*, namely, what would happen if the pedal were not pressed.

In the interventionist theory these three criteria seem to come along together. For example, Woodward comments: "A deeper explanation for the behaviour of the car would need to appeal to [generalizations and] engineering principles that are invariant under a much wider range of changes and interventions. Not coincidentally such a deeper explanation could be used to answer a *much wider* range of what-if-questions" (Woodward 2003, p. 260, my emphasis). I have stressed the expression "not coincidentally" because it lets us see that, according to the interventionist theory, a causal explanation that meets the first two criteria will also meet criterion 3. The following passage also suggests that the three criteria of explanatory depth come along together: "Some generalizations are not invariant under any (testing) interventions at all, and hence are non-explanatory. Other generalizations are invariant under some testing interventions (and answer some what-if-questions), and hence are above the threshold of explanatoriness, although they are less invariant and answer a narrower range of what-if-questions than others, and for this reason are less explanatory (Woodward 2003, p. 369).

A consequence of a given explanation meeting criteria I, II and III, according to the interventionist theory, is that the explanation will be relevant to the manipulation and control of the explained event. And this is precisely what the second explanation of the car's acceleration achieves. And, according to interventionism, this is what causal explanations should aim for. The interventionist theory also defines a minimal condition for explanatory depth:

The minimal condition for successful causal explanation is to present a generalization G (relating the putative cause C to the effect E) and show that there is *at least one intervention* I under which G is invariant. It would be much better if we show that G is invariant under many interventions. But showing only one is sufficient for our explanation to be considered as minimally successful. If a generalization is not invariant under any intervention, then it will fail to qualify as invariant or explanatory. In that case both the generalization G and the putative explanation associated to G fall below what Woodward calls "the threshold of explanatoriness" (see Woodward 2003, p. 203; p. 368).

3. Case study: cholesterol and atherosclerosis

3.1. Detailed causal explanation

Let us analyze the causal relationship between cholesterol and atherosclerosis (also known as arteriosclerotic vascular disease or *ASVD*). Sterols are substances in our body that serve, for example, as material to build cellular plasmatic membranes, or for producing bile. Among all the different types of sterols, two are particular relevant factors for heart diseases: high-density lipoproteins and low-density lipoproteins.

The low-density lipoproteins (*LDL*) are sterols of a very harmful type. *LDL* have low density because they are composed by low amounts of proteins and great amounts of lipids. If the *LDL* are not properly eliminated by the organism, they start accumulating in the walls of the blood vessels, forming atheromas. These atheromas may reduce or even obstruct the flow of blood, avoiding the proper oxygenation of the heart and the brain. Additionally, oxidized *LDL* molecules damage the vessels walls making them weaker. These are the reasons why *LDL* are often referred as "*bad cholesterol*".

"High-Density Lipoproteins" (*HDL*), on the contrary, are molecules highly dense in proteins, and poor in lipids, which help to *prevent* heart diseases. Most of the *HDL* in our body are produced in the liver and they are transformed in bile to be used in digestive processes. *HDL* also travel in the blood, collecting the cholesterol that cells did not use. Those *HDL* arrive to the liver and they are decomposed and recycled for bile production. For these reasons *HDL* are also known as *good cholesterol*.

The algebraic addition of HDL and LDL is defined as *total cholesterol* TC. TC = HDL + LDL. And the following table (3-1) contains the recommendable levels of HDL, LDL and TC respectively (see Tudela 2000, pp. 33-34).

	TC (mg/dl)	LDL (mg/dl)	HDL (mg/dl)
Desirable	<i>TC</i> < 200	100 < LDL <	HDL = 45
concentration		129	
Limit of	200 < TC <	130 < LDL <	The highest the
acceptable	239	150	concentration of
concentration			HDL, the better
High risk	240 < TC	160 < LDL <	
		189	
Very high risk		190 < <i>LDL</i>	

Table 3-1

Chapter One

Using the information we have gathered so far, together with the possible values of the variables in table 3-1, someone could propose the following causal graph as a representation of the causal relationships among good cholesterol *HDC*, bad cholesterol *LDC*, total cholesterol *TC*, and atherosclerosis *ASVD*.



Fig.3-1. Causal Graph 1

Let us now imagine that, by means of the interventionist theory, we want to assess whether the putative causal relationship between the variables TC and ASVD is genuine. In order to do so, we shall firstly propose some possible manipulations to control the value of the postulated causal variable TD. Afterwards, we shall verify which of them fulfil the conditions **IN** (defined in sec 2.2). This will provide a set of *interventions* for testing the causal relationships postulated in the explanations under study. Once we have the set of testing interventions, we can proceed to analyse if the causal link between TC and ASVD remains invariant under the interventions. The higher the invariance under interventions, the deepest the explanation of heart disease caused by cholesterol would be.

How can we manipulate the values of TC? A natural way to start manipulating TC values is altering the ingestion of specific ingredients that have been proven to increase or decrease cholesterol levels. The following table contains some examples (see Tudela 2000, pp. 51-71):

Ingredients	Effects	
1. Soya Milk and Orange	Reduce LDL	
juice		
2. Asparagus and broccoli	The liver spends <i>LDL</i> producing the high	
	amounts of bile required to digest these	
	vegetables	
3. Tomato	Inhibits LDL production	
4. Margarine, eggs, and fat	Increase <i>LDL</i>	
meat		
5. Red wine	Prevents <i>LDL</i> from sticking in the vessel's	
	walls; It also increases HDL	
6. Almonds, peanuts,	They increase HDL for they contain mono-	
pistachios, etc.	saturated lipids. They avoid <i>LDL</i> oxidation.	
	They must be consumed moderately	
7. Avocado	Contains mono-saturated lipids, increases	
	HDL	
8. Cacao	It is a mono-saturated lipid, so it increases	
	HDL. But it must be consumed as dark	
	chocolate, or even better as grain, to avoid	
	factors of obesity as whole milk and sugar	

Table 3-2

Using table 3-2 several manipulations of *TC* can be defined by including different amounts of these ingredients in the patient's diet. All such manipulations will refer to ingested cholesterol (*exogenous cholesterol*).

Additionally, we can propose some manipulations of produced in the liver (*endogenous cholesterol*). The production of endogenous cholesterol is determined by the capacity of the liver's cells to capture and decompose *LDL* molecules. There is a genetic mutation that prevents the liver from fabricating receptors of *LDL* molecules. As a consequence, even if the blood has a high level of cholesterol, the liver cells cannot perceive it, and they produce high amounts of *HDL*, causing, of course, an important race of the total cholesterol in the blood. Let me call this increment in the production of *HDL* due to a genetic modification "the mutation manipulation".

It is worth noting that, in fulfilment of criterion III, our knowledge about cholesterol also allows us to answer several *what-if-questions*: What would happen if the genetic constitution prevents the liver to absorb cholesterol? What would have happened if the patient had not eaten eggs and bacon every morning? And so on. So, intuitively, given the proposed manipulations in endogenous and exogenous cholesterol, and given the answered *what-if-question*, it seems that we have collected a good amount of knowledge to provide a *deep* causal explanation of atherosclerosis due to levels of cholesterol. Let us see if our intuition is correct.

3.2. Interventionist analysis of the detailed causal explanation

In order to prove the depth of our causal explanation, the interventionist theory indicates that firstly we must check that our manipulations fulfil the conditions IN (that is, that they count as *interventions*) and, secondly, we must show that the causal links in the causal graph 1 remain invariant under those interventions.

The first difficulty we find is that the interventions in table 3-2 do not fulfil requirement **IN-i**. According to the requirement **IN-i** if we wiggle the values of *HDL* as an intervention upon *TC*, such intervention should be performed while disrupting the causal relationship between *TC* and any other previous causes (*LDL*). But, in this case, that is impossible as, by definition, HDL = TC - LDL, and this means that changing the values of *HDL* and *TC*, necessarily implies changing the value of *LDL* as well.

This difficulty is however easy to solve by revising the variables selected in the causal graph 1. On the one hand, the causal variable TC is a function of two underlying factors: LDL, which is causal, and HDL, which is actually preventive. That is the reason why the manipulations proposed in table 3-2 turn out to be what Sprites and Scheines (2005) called "ambiguous manipulations". On the other hand, the causal graph 1 seems to represent, wrongly, that LDL causes TC, or HDL causes TC. But the arrows from HDL or from LDL to TC are not causal links, but rather mathematical or definitional links, and this should be expressed in some way (see Woodward 2011, p. 24). A possible way to repair the mistaken assumptions in causal graph 1, is removing TC and keeping only the underlying factors HDC and LDL, and representing the definitional relationship between them by a double arrow (different from the simple arrows that represent causal relationships). The emended causal graph would be the following:



Fig. 3-2. Causal graph 2

Once we accept that the causal graph 2 is correct, we proceed to check again if our manipulations in table 3-2 fulfil the conditions **IN**. But now a second difficulty appears: according to requirement **IN-iv**, the values of *HDL* should remain fixed while we wiggle the values of *LDL*, and vice versa. However, due to the intertwined relationship between *HDL* and *LDL* this seems impossible. Levels of *HDL* and *LDL* may influence each other. For example, high amounts of *LDL* in the blood may inhibit the production of *HDL* in the liver. Or, in a similar but opposite way, if *LDL* levels break down very drastically, the organism may perceive lack of cholesterol in the blood and an increase in the production of *HDL*. This interrelation between *HDL* and *LDL* levels seems to prevent all the dietetic manipulations to meet requirement **IN-iv**.

The only manipulation that meets **IN-iv** is the mutation manipulation because, as the body itself stops perceiving *LDL* levels, the *HDL* production increases *leaving the amount of LDL unaffected*.

The conclusion of interventionist analysis is that all our manipulations of exogenous cholesterol based on diet modifications fail to meet requirement **IN-iv** and thus they do not classify as proper interventions. The only manipulation that counts as a proper intervention is the "mutation manipulation" of endogenous cholesterol. Although the criterion III is widely fulfilled by answering several what-if-questions, the interventionist criteria I and II of explanatory depth are fulfilled only by one single intervention.

The three interventionist criteria of explanatory depth work harmonically in the car's example (in sec.2.4), but this does not occur in the cholesterol-atherosclerosis case. And it seems frustrating to accept that our explanation falls just above the threshold of explanatoriness. We would like to claim that cholesterol levels *due to diet modifications* are also causally related to atherosclerosis, specially taking into account that there is a great amount of *what-if-questions* that can be answered using the list of ingredients in table 3-2. In the next section we will discuss what could possibly go wrong in our interventionist analysis. But, for the sake of the argument, let me first present a possible simplification of the causal explanation of atherosclerosis based on cholesterol levels.

3.3. Simplified causal explanation of atherosclerosis

Imagine an absurd, but still possible, causal explanation of atherosclerosis such as one that ignores all knowledge about *HDL*, and ignores the mutation manipulation, and it is therefore completely based on knowledge about levels of *LDL*. The causal graph 3 bellow represents this of simplified causal explanation of atherosclerosis.

$$LDL \longrightarrow ASVD$$

Fig. 3-3. Causal graph 3

In this imaginary simplified explanation, only a reduced subset of manipulations can be proposed. Out of the dietetic modifications proposed in table 3-2, only the first five can be used in this case. Nevertheless, and surprisingly, this time the interventions actually meet **requirements IN** because there is no other factor to fix while we intervene upon the variable *LDL*. Therefore, the causal relationship illustrated in the causal graph 3 turns out to be genuine, and consequently this simplified explanation fulfils the three interventionist criteria of explanatory depth for a number of interventions. Is it not paradoxical that this explanation is evaluated as *deeper* than the detailed explanations presented before?

4. Discussion

Our analysis shows that the criteria of explanatory depth, as originally defined in interventionism, may lead to counterintuitive results. The result that the simpler explanation, which deliberately ignores a relevant causal factor, is *deeper* than the detailed explanation is clearly wrong, and is a *counterintuitive result from the interventionist perspective itself.* The detailed explanation, which accounts for a wider set of counterfactual questions, should be *deeper* as it happens in the example of the car's acceleration.

At this point one could simply assume that Woodward's interventionism is not applicable to cases with multiple correlated causes. One could thus dismiss interventionism and look for alternative theories of causation. However, there are at least two motivations to avoid this negative reaction and, instead, continue defending interventionism.

Firstly, interventionism is a practical tool to test causal relations and, maybe more importantly, it is able to tell us *how to modify* the scientific explanations in order to get a better *control* over the effects. Interventionism allows detecting *genuine causal links*, and identifying *deeper explanations*. In our case study, for example, the interventionist theory has pointed at the fact that controlling levels of total cholesterol *TC* was not as good a strategy to prevent atherosclerosis, as controlling the underlying factors *LDL* and *HDL*. In other words, interventionism is able to point at the factors that doctors and patients should be measuring, and it guides us to a better understanding of diseases. This fact, *by itself*, would be already worthy to continue using the interventionist theory. Besides, given that the majority of diseases have multiple causes that influence each other's occurrence in several ways (see van Loo et al. 2012), having a theory of causation for assessment of this kind of cases is necessary.

Secondly, Woodward (2003, p. 132; 2011) has a flexible attitude towards possible changes and improvements of interventionism. The challenge for interventionism brought about by our case study is whether the interventionist theory is capable of telling us something about the causal link between a given effect E and its putative causes C_1 , C_2 , C_n , when those causes are correlated to each other. Even in those cases, Woodward believes that we can legitimately use counterfactuals to elucidate causal claims along the lines suggested in the interventionist theory.In cases like this, Woodward (2003, sec. 5.11) suggests to verify whether two conditions-indeed two original motivations for introducing the notion of intervention- are being met or not: Firstly, we must ensure that "there is a basis for claims about what will happen to E under an intervention on C"; i.e. we should be able to associate some well-defined notion of change with C, and we have some grounds for saying what the effect, if any, on E would be of changing just C and nothing else. And secondly, there must be "a way of disentangling the effect on E of changing just C from the effects on E of changes in other potentially confounding variables" (see Woodward 2003, pp. 131-2). In a more recent paper (2011), Woodward also shows his disposition to modify and improve interventionism. This has been a strong incentive to propose here an attempt to make interventionism more suitable for cases where it seems difficult to apply it.

Chapter One

My claim is that, re-defining and understanding in a new way the interventionist criteria of explanatory depth, the counterintuitive results of the kind we obtained in section 4.1, can be avoided. This proposal is meant to *complement* Woodward's efforts to improve interventionism. The new interventionist vectorial notion of explanatory depth, which will be developed in the next section, aims to make the interventionist theory applicable in a more natural way to multi-causal explanations in the social, natural and health sciences.

5. Proposal: Re-defining notion of explanatory depth

The aim of this section is to propose a new way of understanding the interventionist notion of explanatory depth that results more suitable for multi-causal explanations. The key conceptual modification of the notion of explanatory depth, embedded in the new vectorial proposal, consists in assigning much more importance to counterfactual questions. In Woodward's original proposal, invariance under intervention was primordial. In the proposal here presented, counterfactual answers are almost as important as invariance under intervention. There is a single sense in which interventions continue being more important than counterfactual answers: in order to cross the explanatoriness threshold, invariance under at least one intervention is still required.³ However, once the threshold has been passed, the amount of counterfactual answers is as significant as the amount of possible interventions under which the causal link postulated in the explanation remains invariant. In order to test the explanatory depth of a given causal explanation, following steps is recommended:

 1^{st} : Find a variety of manipulations of the values of the putative causes, and verify if those proposed manipulations fulfil requirements **IN**. If, for any reason, the manipulations fail to meet the requirements, modifications in the relevant set of variables, possible definitional relations among the variables, and consequent modifications in the causal graph must be considered.

 2^{nd} : Once the relevant variables, the causal graph, and the manipulations upon the putative causes are chosen, it must be ensured that the generalization figuring in the explanations remains invariant under *at least one* intervention, crossing the threshold of explanatoriness.

 3^{rd} : Build a vector whose component X is the number of *different kinds* of manipulations upon the causes, whose component Y is the *number of proper interventions* under which the generalization remains invariant, and

whose component Z is the number of *what-if questions* answered. In other words, each of the original criteria of explanatory depth I, II and III is a component of this vector, which we will name "the interventionist vector of explanatory depth".

Applying this vectorial notion of explanatory depth, we not longer arrive to the counterintuitive result that the simplified causal explanation of atherosclerosis (sec 3.3) is *deeper* than the detailed explanation (sec 3.1). The mutation manipulation places both explanations above the threshold of explanatoriness. However, the detailed explanation presents three different types of manipulations (endogenous cholesterol, exogenous cholesterol and mutation), while the simple explanations can only make use of the last two. Furthermore, the fact that the detailed explanation is able to answer much more *what-if-questions* makes the magnitude of its explanatory vector much larger than the vector corresponding to the simplified explanation. Therefore, the vectorial notion of explanatory depth leads us to a correct causal assessment of the two explanations of atherosclerosis, i.e., it leads us to the conclusion that the detailed explanation is *deeper* than the simplified one.

The vectorial notion proposed in this paper may seem, at first sight, a simple change from verbal enunciation to mathematical representation of the criteria. However, the new vectorial notion *essentially affects* the notion of explanatory depth for it replaces necessity by sufficiency regarding criterion II and criterion III.

6. Conclusions

Interventionism not only provides an innovative methodology to test causal explanations, but also helps to improve the causal explanations already existent in different areas of knowledge, and has a practical *pay off* when it elicits how to manipulate the effects that are relevant for different reasons. As an example of this, in section 3.2, interventionism allowed us to identify that a causal explanation of atherosclerosis based on *HLD* and *LDL* levels, was more convenient than a causal explanation based on the total cholesterol (*TC*). Therefore, interventionism provides us with specific variations and proposals to improve causal explanations. This is already a good reason to defend Woodward's interventionist theory.

According to the original definition of *explanatory depth*, the detailed explanation would be just above the threshold of explanatoriness. The simplified explanation that ignores a lot of relevant factors of atherosclerosis

would be considered *deeper* than the detailed explanation. And this is a counterintuitive and incorrect result.

The advantage of the re-definition of the interventionist notion of explanatory depth, here proposed, is that it allows us to understand why some explanations are deeper than others, even though the number of interventions is reduced due to the correlations among different causes.

This vectorial notion of explanatory depth aims to be particularly fruitful for cases in health sciences, where the majority of diseases present complex multi-causal patterns. Nevertheless, it also aims to be suitable for multi-factorial causal explanations provided in all kind of scientific practices including social sciences and foundations of physics (see Samaniego 2013). This proposal supports the lines of defending and improving interventionism, and hopefully, helping scientists to reach a better understanding and control of relevant causal factors.⁴

References

Baumgartner, M. and L. Glynn. 2013, "Introduction to Special Issue on 'Actual Causation", *Erkenntnis*,

http://link.springer.com/article/10.1007/s10670-013-9441-8

- Haavelmo, T. 1944, "The Probability Approach in Econometrics", *Econometrica* 12, 1
- Marcellesi, A. 2010, "L'interventionnisme permet–il la causalité "descendante"?" *Igitur, Arguments philosophiques* 2 (1), http://www.igitur.org/L-interventionnisme-permet-il-la
- Menzies, P. and H. Price 1993, "Causation as a Secondary Quality", British Journal for the Philosophy of Science 44, 187
- Phyllis, I., F. Russo and J. Williamson (eds.) 2011, *Causality in the Sciences*, Oxford: Oxford University Press
- Reutlinger, A. 2013, A Theory of Causation in the Social and Biological Sciences, United Kingdom: Palgrave Mcmillan.
- Russo, F. 2012, "Explaining Causal Modelling. Or, What a Causal Model Ought to Explain" in M. D'Agostino *et al* (eds.) *New Essays in Logic and Philosophy of Science*, London: College Publications
- Samaniego, F. 2013, "Causality and Intervention in the Spin-Echo Experiments", *Theoria* 21 (3), 477
- Spirtes, P. and R. Scheines 2005, "Causal Inference of Ambiguous Manipulations", *Philosophy of Science* 71, 833
- Suárez, M. and I. San Pedro 2011, "Causal Markov, Robustness and the Quantum Correlations" in *Probabilities, Causes and Propensities in Physics*, Synthese Library 347, 173, Berlin and New York: Springer