Causation & Mental Causation

A contemporary debate on mental causation issues informed by different perspectives on causation

... Do not all charms fly

At the mere touch of cold philosophy?

There was an awful rainbow once in heaven:

We know her woof, her texture; she is given

In the dull catalogue of common things.

Philosophy will clip an Angel's wings,

Conquer all mysteries by rule and line,

Empty the haunted air, and gnomed mine—

Unweave a rainbow, as it erewhile made

The tender-person'd Lamia melt into a shade.

– John Keats—Lamia (1819)

lines 229-238

Introduction

These poetry lines are probably some of the most famous to be written about science and philosophy, and unfortunately do not give a positive picture of them. The supernatural half-serpent being Lamia is transformed by Hermes from a snake into a beautiful woman; she then seduces prince Lycius, but at their wedding feast, Apollonios of Tyana, a neopythagorian philosopher, sees Lamia's true nature and breaks the illusion of her beauty. In a parallel way, according to Keats' poem, when the rainbow was explained by science, when "its true nature is seen", it lost its beauty and mysterious power.

The rainbow, however, is still there and it reveals its beauty despite being explained. It is the task and ambition of philosophy and science to explain things in the world, to reveal its inner sturcture. Scientific and philosophical explanations constantly ask the question "Why" and search the world for satisfying answers. From this simple question, which humans some tens of thousands of years ago, started asking themselves, a whole civilization began. To take this question seriously means to begin realizing that certain things cause other things. No other species grasps this to the extent that we do.

One of the most difficult tasks of both philosophy and science is the explanation of ourselves as beings, since we then automatically become both the explorers and the explored ones. How do our minds function; what are our thoughts consisted of; and in virtue of what do we decide how to act? These questions are part of philosophy of mind, which studies the nature of minds and their relationship to the physical world. Questions concerning causal relations, causal inference and their metaphysics, are topics in the philosophy of science. In the present thesis a combination of these two branches of philosophy will be attempted.

In the next section, the contemporary issues in philosophy of mind will be outlined and a map of mental causation theories will be sketched. After this, the discussion will focus on the most prominent contemporary causal theories, starting from Hume and continue with his legacy; 20th century causal theories, such as Probabilistic Causation and Counterfactual Causation. Anti-Humean approaches to causation, which consider causal relations as an intrinsic mechanism between the relata, will be the focus on the second half of the same section. Finally, in the last

two sections the longed-for combination of philosophy of mind and philosophy of science will be attempted. Issues discussed in the first section on mental causation will be informed by the different approaches on causation, shared in the second section. In this way, the overall aim is to "connect the dots" between distinct, but interconnected philosophical branches and offer a broader perspective on both mental causation issues and causation in general.

Mental causation: An Overview

The problem of mental causation consists of the philosophical exploration of the causal interaction between the mental and the physical world. There are parts of this world, which we call mental (such as intentions, desires, decisions, thoughts) and we believe that they have the ability to cause physical events (such as bodily movements). But how and in virtue of what does this interaction take place?

In the history of philosophy this problem was initially captured as the mind-body problem, the relation between an immaterial mind and the material body. This discussion has already taken place in the antiquity, but it was Descartes in his *Meditations* (Descartes, 1641) that sealed the modern discussions on the topic.

In his *Second Meditation* (Descartes, 1641), he asked '*But what then am I?'* to which he replied '*A thing which thinks'*, and concluded in the sixth Meditation, that '*it is certain that this I is entirely and absolutely distinct from my body, and can exist without it'*. Descartes's view was that minds and bodies constitute two distinct categories of substance: minds are immaterial substances whose essential nature is thinking, while bodies are material substances located in physical space whose essence consists in being extended in space.

Substance dualism of this form was common at the time. However, Descartes, urged a further view: minds and bodies are in causal interaction with each other, minds influencing bodies in voluntary actions and bodies influencing minds in perception and sensation. The problem of mental causation was born out of this question regarding the causal interaction between two ontologically distinct substances. How can substances with such diverse natures, one with thinking as its essence

and not located in space and the other with extension and location in space, causally interact with each other?

Descartes conception of the body as something that does not move by itself, but only when it comes into contact with other objects or substances suggests that physical causation requires contact—the object that causally influences another must come into contact with it (see (Jaegwon, 2009)). The very idea of contact, however, requires spatiality.

This problem was captured in Princess Elisabeth's letter of May 1643 (Bennett, 2017) to Descartes, in which she asked him to explain:

"how the mind of a human being can determine the bodily spirits in producing voluntary actions, being only a thinking substance. For it appears that all determination of movement is produced by the pushing of the thing being moved, by the manner in which it is pushed by that which moves it, or else by the qualification and figure of the surface of the latter. Contact is required for the first two conditions, and extension for the third. [But] you entirely exclude the latter from the notion you have of the soul, and the former seems incompatible with an immaterial thing." (Bennett, 2017 p. 4)

Descartes's reply was that mind-body causation must be understood through the idea of mind-body union, and that this is a primitive notion that is intelligible per se. Since there is unity between mind and body, the issue of contact does not arise with respect to it. But Princess Elisabeth unsatisfied by Descarte's response writes back:

"And I admit that it would be easier for me to concede matter and extension to the mind than it would be for me to concede the capacity to move a body and be moved by one to an immaterial thing" (Bennett, 2017 p. 5)

This is the first known causal argument for physicalism. Elizabeth actually claims that the mind should rather be physicalized than ground causal relations between immaterial minds and material bodies.

The difficulties with mental causation within the Cartesian setting depend on a particular assumption; namely, that physical causation requires contact between cause and effect. It can be argued, however, that causation is in general inseparable from spatiality, and thus all causal relations involving immaterial substances are equally problematic with the original mind-body causation.

What is eminently relevant in Descartes's substance dualism is his theory of interaction between body and mind, while most of his other contemporaries, such as Leibniz's monads, Malebranche's occasionalism or Spinoza's parallelism, denied any mind-body interaction. It is his dualism that brings about the difficulties in his interactionism and has been left behind.

Contemporary Mental Causation Problems

Questions about the interaction between the mental and the physical and about the existence and nature of mental causation are prominent in contemporary discussions of the mind, as well. As already stated, originally the problem of mental causation was that of understanding how an immaterial mind could interact with the body. Most philosophers nowadays deny the existence of immaterial souls, but the problem of mental causation has not gone away. Instead, focus has shifted to mental properties. Thus, the contemporary problem of mental causation lies in its commitment to property dualism. How could mental properties be causally relevant qua being mental?

The re-emergence of mental causation as a major problematic in the 1960s was occasioned by the reluctance on the part of most philosophers of mind to go all the way with physicalism, which dominated contemporary discussions. These philosophers reject substance dualism, recognizing only material things as the inhabitants of the world, and yet they are committed to the preservation of mental properties.

Davidson's Theory of Mind

One of these philosophers, who changed the agenda on mental causation with the publication of his '*Mental Events'* (Davidson, 1970), is Donald Davidson.

Event Ontology and Event Causation

Davidson's entire theory of mind is based on his event ontology, one of the most important and distinctive feature of his theory.

Events are entities, as opposed to processes or states (Davidson, 2001). Events enjoy a kind of ontological primacy and it is also important to note that for Davidson, properties are not ontological constituents of events. Cause and effect hold only between events and not in virtue of the properties (or descriptions) of those events (Davidson, 1995). Since events themselves stand in causal relations, causality holds between events no matter how those events are described.

Events can be described in differing vocabularies and therefore make the ideal subject matter when contrasting or identifying the mental and the physical. Davidson claims that the physical realm is "recessive" in relation to the mental: every mental event is necessarily physical, but not every physical event is necessarily mental (Davidson, 2001 pp. 211, 214). Mental events are distinguished by their intentionality, and while all mental events are identical with certain physical events, there are some physical events that do not exhibit intensionality—they are explained in purely extensional terms.

As already pointed out, according to Davidson's event ontology, causal relations can hold only between events (Davidson's "event causation") and they obtain no matter how the events are described, so that sentences ascribing them are extensional. Ascriptions of causal relations need not therefore, explain phenomena. For example, saying that what I just referred to was the event-cause of something that happened some months ago, does not explain what happened some months ago.

If events are causally related, there must be a strict law connecting them. "Where there is causality, there must be a law: events related as cause and effect fall under strict deterministic laws" (Davidson, 2001 p. 208). Strict laws are strict only if the events described belong to a closed system, where a closed system is such a system that whatever can affect the system is part of the system being described. Only physics is such a system, so causally related events follow a law of physics. Causality as a relation between events is a physical, nomological relation (Davidson, 2005).

Event causation also involves mental events. But, according to Davidson: "The efficacy of an event cannot depend on how the event is described, while whether an event can be called mental, or can be said to fall under a law, depends entirely on how the event can be described" (Davidson, 2005 p. 191). So, mental events are events that have a mental description. But since events are causally related no matter how they are described, mental events can be causally related to either physical or mental events.

An explanation, in contrast to a causal relation, relates to sentences (propositions, facts) and not to events. The point of explanation is to render phenomena intelligible, and what does so under the description of the phenomena may not do so under another. One and only phenomenon may have different kinds of explanation, each explaining it under a different description, whereas there is only one way in which it is causally related to its cause or its effect.

Strawson also discusses this point of Davidson. He writes: "*Causality is a natural relation that holds in the world between particular events or circumstances, just as the relation of temporal succession does or that of spatial proximity… But if causality is a relation which holds in the natural world, explanation is a different matter… It is an intellectual or rational or intentional relation and does not hold between things in the natural world… [but] between facts or truths*." (Strawson, 1985)

Davidson's ontology of action is "*ontological monism accompanied by an uneliminable dualism of conceptual apparatus… There is only one [kind of] substance [but] the mental and the physical are irreducibly different modes of apprehending, describing, and explaining what happens in nature*." (Davidson, 2005 p. 290)

Anomalous Monism and The Supervenience Thesis

The question however about the ontological relations between the sphere of the mental and the sphere of the physical still remains. If intentional actions entail mental events, then what is their exact ontological relation to physical events? Davidson argues that, acting because of a certain reason entails that there was an event associated with that reason that caused the action (Davidson, 2004 p. 109). However, a mere claim of association is not an ontological thesis. What is the ontological connection between this associated event and the mental state? A closer look at Davidson's Anomalous Monism will hopefully elucidate these open questions.

By developing Anomalous Monism, Davidson aims to resolve a contradiction in three principles that he holds to be true of mental events:

(1) The Principle of Causal Interaction: At least some mental events causally interact with physical events.

(2) The Nomological Character of Causality: Events that are causally related fall under strict deterministic laws.

(3) The Anomalism of the Mental: Mental events cannot be explained or predicted by strict deterministic laws.

Davidson reconciles these three apparently contradictory principles by arguing that the three in conjunction are compatible after all, for they imply monism, i.e., that all mental events are identical to physical events (Davidson, 2001 p. 209). If this is the case, then mental and physical events can be in causal relation and causal relations can require laws. However, there need not be mental-physical laws because strict lawlike relations can hold between certain physical and mental events—mental events that are identical with certain physical events.

The identity of mental and physical events that allows all three principles to be compatible is token identity: while mental properties or types cannot be strictly identified with physical properties or types, specific mental particulars can be identified with specific physical particulars; i.e., they are event-tokens. Although mental properties are irreducible to physical ones, mental events are physical events.

In order to preserve Anomalous Monism's claim to physicalism, Davidson holds that there is a supervenient relation between the mental and the physical. Specifically, while mental and physical tokens are identical, mental types are supervenient or dependent on physical types. In this way Davidson hopes to find a middle position between type-identity and property dualism. Through the supervenience thesis, Davidson wants to avoid the existence of strict laws connecting the mental and physical realms in order to preserve what he holds to be true of the mental—that the mental cannot be explained or predicted by strict psychophysical laws (Davidson, 2001 p. 208). This kind of nonreductive physicalism is according to Davidson, compatible with his Anomalous Monism. Anomalism of the mental does not imply that there can be no psychophysical laws, but rather that there can be no strict psychophysical laws.

One could claim, however, that if we more fully investigate the non-strict laws linking the mental and the physical we would get a more robust connection between the two and this would challenge Davidson's rejection of strict psychophysical laws. But Davidson thinks that even if we could pick out each mental event using only physical predicates, the mental would not be reducible to the physical or defined within it. The mental and the physical pick out the same event (they have the same extension by token-identity), but they pick it out in different ways (they produce different descriptions of the same event). A shift in vocabulary is necessary. Thus, the mental retains its

anomalism insofar as mental types are not reducible to physical types (and therefore Davidson need not be committed to strict correlating laws). But if singular mental events (tokens) can be identified with physical tokens, then mental events can causally interact with physical events. Because they are described in a physical vocabulary, mental events are physical events also, and physical events stand in causal relation to each other. Thus, Anomalous Monism supplemented by supervenience offers Davidson a robust token-identity form of physicalism.

However, Davidson's opponents have argued that this solution, Anomalous Monism combined with supervenience, leads to epiphenomenalism. So we now consider the accusation of epiphenomenalism and Davidson's response.

The Problem of Epiphenomenalism

Epiphenomenalism is a position in the philosophy of mind according to which mental states or events are caused by physical states or events in the brain but do not themselves cause anything. It states that the mental is just a by-product, it is causally inert. The charge of epiphenomenalism arises because of the conclusion of token identity by Davidson. To sum-up, causation requires strict laws and there are no such mental-physical laws, yet mental events cause physical events. This is because mental events are physical events. Then all the causal work that a mental event does is due to its physical properties. Hence, the mental may be relevant to the physical, but the mental is causally inert because an event's causal powers rest entirely on its physical properties.

In his essay "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?" (Kim, 1995) accepts Davidson's notion of causality only to push the question further: given Anomalous Monism's distinction between causality and causal explanation, it seems that the mental cannot causally explain physical events. The issue, holds Kim, is and "has always been the causal efficacy of properties of events" (Kim, 1995 p. 224). Kim indicates that Davidson is denying the mental any role in explanation, i.e., it is not in virtue of its having a certain mental property that an event causes anything. For example, if mental event m1 is identical to physical event p1 (they are tokens of the same event e1) and p1 causes p2, then m1 causes p2, but the mental event m1 (that is, the mental description of event e1) cannot offer a causal explanation of p2. In other words, if events with mental properties have no role in causally explaining events with physical properties, then the

former set of properties is inert with respect to the latter set of properties and is therefore epiphenomenal (Kim, 1995 pp. 21-22).

Davidson has two possible ways of responding in order to maintain the explanatory potency of the mental. Firstly, to argue that the mental is supervenient on the physical means that if there is a mental difference in an event, then there will also be a physical difference, and in turn a difference in the causal relations that the event enters into insofar as it is these physical aspects of the event that causal relations depend on. Davidson argues that since a mental difference in the event changes the relations it enters into, the mental has a role to play in the causal explanation of the event in question.

Davidson's second option is to deny strict psychophysical laws and affirm non-strict or less strict psychophysical laws. For Davidson, strict laws are most likely to be found in a more fully developed physics; they are exceptionless and deterministic as nature (Davidson, 1995 p. 8). For Davidson, strict laws always underwrite singular causal relations. Yet these types of strict laws are rare, and in fact, a great deal of our practical scientific knowledge is non-strict (Davidson, 1995 p. 9). However, given reasonable assumptions, these examples of scientific inquiry may prove to be completely reducible in a more fully developed physics. The mental, on the other hand, can be understood and predicted only in light of other mental events. This is because the mental is holistic and guided by an a priori constitutive principle of rationality, unlike the physical realm, which requires no such rational guiding principle. Thus the realm of the mental is connected with the realm of the physical through non-strict laws. Yet if all that is required for causal explanation is non-strict laws, which seems to be the case if also other practical sciences, such as geology for example, offer causal explanations, then the mental is explanatorily potent.

According to Kim, however, Davidson's appeal to supervenience only guarantees the "causal relevance" of mental properties and not their "causal efficacy" (Kim, 1995 pp. 23-24). Mental properties are "relevant" to causal explanations since what mental properties an event has affects what physical properties an event has, and it is the physical properties of events that causally explain other events.

And as for Davidson's second argument, his affirmation of non-strict or less strict psychophysical laws, Kim warns advocates of Anomalous Monism not to embrace them: "Where there is a non-strict psycho-physical law, there must be a strict psycho-physical law waiting to be

discovered" (Kim, 1995 p. 24). This is problematic for Davidson's account, since if non-strict laws connect the mental and the physical, then there are some yet-to-be-discovered strict psychophysical laws, exactly what Anomalous Monism denies (Kim, 1995 p. 24). However, Kim seems to assume that irreducible non-strict laws are impossible.

In response, Davidson argues that strict laws do not underwrite the non-strict laws linking mental-physical events. Non-strict psychophysical laws are not simply hidden strict laws – they are valid and independent law-like relations. Non-strict laws are not couched in the vocabulary of a closed comprehensive system, as strict physical laws do. They most often relate events in two disparate vocabularies (Davidson uses here the notion of heteronomic laws). Davidson holds that with respect to these laws one cannot exhaustively specify all the conditions in advance. The presence of ceteris paribus clauses highlights the fact that non-strict laws are not exceptionless and cannot be made perfectly strict. In addition, the nature of mental states is holistic. It is not that one desire is correlated with just an array of different behaviors, but also that any given behavior to which the desire connects is similarly complex.

In his essay, "You Can Fool Some of the People All of the Time, Everything Else Being Equal: Hedged Laws and Psychological Explanations", (Fodor, 1991) criticizes Davidson's account of the irreducibility of the mental to the physical and holds that if one variety of ceteris paribus law is reducible to strict law, then all ceteris paribus laws are reducible and vice-versa. Fodor's question is, if non-strict laws are different from accidental relations (as they must be), then are they not simply underwritten by strict laws? If so, then we should expect (say, when neuroscience progresses) to find some strict laws forming the base of non-strict mental-physical laws, contra Anomalous Monism. Fodor argues that there is good reason to believe that ceteris paribus psychological laws have genuine truth conditions and that they are no different in kind to those found in other special sciences like meteorology, geology, and biology. Fodor's account conflicts, however, with Davidson's thesis on the anomalism of the mental. Thus, Davidson has to reject it. And in order to reject it, he needs to argue that the ceteris paribus laws of psychology and the ceteris paribus laws of other special sciences are different in kind.

Davidson's argument is that a mental state can be identified only in terms of its relation to other mental states. However, Fodor blocks this path of defense by showing that all special sciences operate within their respective network of laws.

The most important defensive option for Davidson is his appeal at the constitutive ideal of rationality. In order to ascribe a propositional attitude to another person, the ascribed mental states must make rational sense in relation to the person's other mental states and behaviors. The same does not hold for the realm of the physical: we do not rationally assign beliefs and desires to rocks to account for their behavior. Thus, to some degree the constitutive principle of rationality shows that psychology is in principle different from geology and other special sciences. Nonetheless, if psychological laws are sui generis in their non-strictness, then psychological laws are not like special science laws, so we cannot understand them, or their causal explanatory power, on the model of the special sciences.

Physicalism: Reductive or non-reductive?

After the exchanges of Davidson and his critiques on this charge of epiphenomenalism, discussions of mental causation have largely moved away from Davidson's theory. The discussion following Davidson, however, helped to bring the general issue of the causal status of mental properties in focus. Given the physical world and its causal closure, how is mental causation possible?

The principle of the completeness of physics tells us, that for any physical change there is a physical cause insofar as there is a cause at all (for a detailed argument, see (Papineau, 2002)). Hence, mental causes cannot cause anything that is not at the same time caused by physical causes as well.

Emergentism

However, some philosophers have rejected the physical causal closure thesis. Emergentism is a form of property dualism that does exactly that. Emergentists believe that emergent phenomena, like consciousness or intentionality, emerge from their 'basal' physical conditions and can then causally affect downwardly their own physical bases and influence the course of events at the lower level (Sperry, 1969). This is in clear violation of the causal closure of the physical domain, since emergentism considers emergent phenomena to be irreducible to and distinct from the physical phenomena from which they emerge. Whether emergentism is a coherent position is a question that goes beyond the scope of the present paper.

Multiple Realization Thesis

Another popular attempt to "save" mental causation from reductivism is the multiple realizability theory, which asserts that a single mental kind (property, state, event) can be realized by many distinct physical kinds. It is an anti-reductionist and anti-identity thesis and it served as the most influential argument against theories that identified mental states with brain states (mind-brain identity theories). But what does realization actually mean?

Suppose we have a family of interconnected macro-properties (mental properties) of a given system (say a person). Suppose that corresponding to each of these macro properties there is a micro property in this system, and that the family of interconnected micro properties provides a mechanism for explaining the connections among members of the macro family. Then the micro properties realize the macro properties.

Fodor and Putnam initiated the multiple realizability thesis by noting the analogy between computational states and mental states (Fodor, 1965) (Putnam, 1965). Any computational property can be "realized" or "implemented" in a variety of ways (electronic, mechanical etc.), so it would be a mistake to identify any computational property with, say, an electronic property, since the same computational property can be implemented without the electronic property, for example mechanically. If mental properties are computational or functional as well, then we cannot identify them with any neural state; for they could also be implemented non-neurally, e.g. electronically or else (Block, 1997).

Multiple realizability often applies within and between individual people. For example, the central nervous system is often supposed to be quite plastic, especially in young ages. If the brain is injured, knocking out some capacities, the capacities often reappear with different neural realizations. Recently, (Sur, 2003) and his colleagues at MIT rewired the visual system of one eye of neonatal ferretts to feed to the auditory cortex. They found that ocular dominance columns of the sort found in visual cortex formed in auditory cortex and the auditory cortex functioned much like the visual cortex normaly functions. Further, there are many states and capacities that are known to be implemented differently in different people, e.g. the capacity to read.

However, Kim's influential exclusion argument attempts to demonstrate the inconsistency of non-reductive approaches in the philosophy of mind. As we already saw in the discussion of Davidson, the problem with the supervenience thesis, given the principle of physical causal closure and that mental properties supervene on physical properties, mental properties and events cannot be causally efficacious qua being mental. But if they are causally efficacious qua being physical, then – according to Kim – they are either epiphenomenal or they are reducible to their physical equivalences (Kim, 2009). The same argument applies for all aforementioned non-reductive approaches. We will elaborate on the exclusion argument in the final section.

Up to this point, I have tried to roughly sketch out the major problems of mental causation. Since the final purpose of this work is to investigate whether the causal theory we choose plays a role to our argumentative and analytic strategies concerning these problems, we will go on by presenting the major contemporary causal theories. A discussion of their metaphysical and epistemological stand will inform our final discussion on their application to mental causation problems.

Causal Theories

Basic Intuitions & Divisions

There are two central intuitions about causation that need to be taken into account. On these two central intuitions is the basic distinction between Humean and non-Humean accounts of causation based.

Humean-based theories of causation are generally based on regularities in nature; they are theories of general causation. Causal relations are based on the extrinsic factor of regularities. Whether or not a sequence of two distinct events c and e is causal depends on whether or not events like c are regularly followed by events like e. This intuition is underpinned by an epistemic consideration; namely, that we consider a sequence of events c and e causal, only if a regular association between events like c and events like e has been observed.

Non-Humean, on the other hand, ground causation on the intrinsic features of the causal relata, so they are mostly singular. Whether or not a sequence of two distinct events c and e is causal depends wholly on the events c and e and their own properties and relations, that is, it depends wholly on the intrinsic and local features of the actual sequence of event (Psillos, 2002 pp. 6-7).

Now, these two intuitions pull in contrary directions. The regularity intuition implies that a sequence of events is causal if and only if it instantiates a regularity. Hence, it implies that the truth-condition of the relation of cause and effect is extrinsic to its relata. It makes causation dependent on general facts: on what happens at other places and at other times. The intrinsic-relation intuition opposes all this. It takes causation to be wholly dependent on singular facts, on what happens there and then, in the actual sequence of events, independently of any regularities.

It would be an almost impossible task to advance a theory that respects both of these intuitions. Most typically, Humeans base their theories on the regularity intuition, while non-Humeans base heirs on the intrinsic-relation one. However, recent sophisticated accounts of causality try to grasp both the generality and the singularity of causal relations. For example, there are humean-inspired counterfactual theories, which are also singularist.

For this reason, there is a more recent division of theories of causation into two other broad families, which fits best for the contemporary debate on causation; the causation-as-dependence (or causation as difference-making) family and the causation-as-production family. The former includes the regularity view and various counterfactual approaches, including the recent sophisticated account of interventionism. Examples of the latter are Salmon's mark transmission account and Dowe's conserved quantity account, as well as other mechanistic accounts. Hall (Hall, 2004) is a standard reference for this division and he suggests that these two families make up two independent and equally plausible concepts of causation. This distinction, though, is now mostly used to classify competing theories of causation, without such compatibility or pluralistic intentions.

Regularity Theories

Hume's theory of causation

Until Hume philosophers thought there were necessary connections to be found in nature. They thought that, when c causes e, there is something in virtue of which, c brings about or necessitates e. The cause is thought of something that has the power to produce the effect and the effect follows with necessity from the cause.

However, on the received reading of Hume's Treatise (Hume, 1739), there is no place for necessity in nature. Hume, in his famous dictum, writes: "Necessity is something that exists in the mind, not in objects" (Hume, 1739 p. 165). Accordingly, Hume has been typically read as "the great denier of necessary connections" (Lewis, 1986).

According to Hume's regularity theory, causation reduces to spatiotemporal contiguity, succession and constant conjunction (regularity). It reduces, that is, to non-causal facts (Psillos, 2002 p. 19). There is no necessary connection between the cause c and the effect e that goes beyond these three factors.

Hume's regularity theory of causation goes hand in hand with his empiricist epistemology. The cornerstone of this epistemology is the thought that "*all our ideas, or weak perceptions, are derived*

from our impressions, or strong perceptions, and that we can never think of any thing we have not seen without us, or felt in our own minds" (Hume, 1739).

But the idea of causation is not derived from the impression of the properties of an object in the world. Being a cause is not a particular quality of an object. To say that c is a cause is simply a way to describe c in relation to an effect e and not a way to ascribe a property to c. It follows that the idea of causation "must be derived from some relation among objects" (Hume, 1739 p. 75).

As we already said before, the three characteristics of the relation between objects, which make up our idea of causation according to Hume, are spatial contiguity, temporal succession and constant conjunction. Spatial contiguity and temporal succession of objects or events can be directly derived from our sense impressions. But contiguity and succession "are not sufficient to make us pronounce any two objects to be cause and effect, unless we perceive, that these two relations are preserv'd in several instances" (Hume, 1739 p. 87) For, "an object may be contiguous and prior to another, without being consider'd as its cause" (Hume, 1739 p. 77) When we restrict ourselves to singular sequences, there is indeed nothing beyond contiguity and succession to be discovered. There is nothing that can distinguish between a causal and a coincidental sequence of events. Contiguity and succession are characteristics of single sequences.

But as we already pointed out, Hume has offered an account of general causation, based on regularities. Single sequences of events are not enough to exhaust the idea of causation. Hume acknowledges that what is taken to distinguish between causal sequences and coincidental ones is that only the former involve the *idea* of necessary connection between events c and e. But where does this idea of necessary connection stem from, if not from our impression of the world?

To ground his causal theory, he needs to explain the possession of this idea of necessity in a way compatible to his own epistemology. He grounds the idea of the necessary connection between the cause and the effect, which is central to our idea of causation, not as a relation in the world, in the external objects, as we already said, but also not as an idea derived by a single sense impression (Psillos, 2002 p. 27). It is constant conjunction, which is involved in the creation of the idea of necessity. Constant conjunction, in contrast to the ideas of contiguity and succession, is a relation among several sequences, not among a single one. Ascriptions of causation can be made only when a certain sequence instantiates a constant conjunction.

The idea of a necessary connection arises via a basic psychological inferential procedure by which the observed past constant co-occurrence of Cs and Es leads us to believe that upon a new perception of a C, an E will necessarily follow. The more frequently we observe Cs being followed by Es, the stronger the belief that, upon a new c, an e will follow (Hume, 1748 p. 58). And he claims that this is the product of a psychological procedure, in particular of the "custom to transfer the past to the future". He grounds thus the idea of necessity on a psychological process inside the individual.

Although causation is subject to philosophical analysis, it is also a "natural relation", since it is the way in which the mind operates. This is an empirical claim, but for Hume it is an important one because "'tis only so far as [causation] is a natural relation, and produces a union among our ideas, that we are able to reason upon it, or draw any inference from it" (Hume, 1739 p. 94). In other words, causation might be the object of our analysis, but it is also presupposed for the functioning of the mind. By looking at causation as a natural relation, he discovers the origin of the idea of necessary connection.

On Hume's theory, the customary transfer of the past to the future happens automatically and unconsciously. The belief that e will happen given c is as "unavoidable as to feel the passion of love, when we receive benefits; or hatred, when we meet with injuries" (Hume, 1748 p. 46). So, concerning what happens in the mind, Hume's point is that what has appeared to be an inference is nothing but a "customary transition" (Hume, 1739 p. 103) from a certain impression or memory to an idea of a necessary connection. Hume stresses that "the custom operates before we have time for reflection" (Hume, 1739 p. 104). And custom, according to Hume, is "a principle of human nature" ; it is "the ultimate principle, which we can assign, of all of our conclusions from experience" (Hume, 1748 p. 43).

But does that mean that a sequence is causal (if and) only if it instantiates a constant conjunction among the relevant event types? And also how do we know if the idea of a necessary connection depends on the inference as Hume thought or if the causal inference depends on the necessary connection?

Hume argues that "the necessary connexion depends on the inference" (Hume, 1739 p. 88). But as we said, it's not a rational inference; it is the "customary transition", which makes us infer the effect from its cause.

To sum up Hume's argument, the idea of constant conjunction does not arise from any new impression in the objects, yet it is the source of the further idea of necessary connection. Hume has unraveled the "essence of necessity" and has found that it "is something that exists in the mind, not in objects" (Hume, 1739 p. 156). Power and necessity "are consequently qualities of perceptions, not of objects, and are internally felt by the soul, and not perceiv'd externally in bodies" (Hume, 1739 p. 166).

What needs to be stressed is that an individual sequence of events is deemed causal only because something extrinsic to the sequence occurs, be it the constant conjunction of similar events, or the customary transition of the mind from the appearance of the one, to the idea of the other. This is why Hume's account of causality is extrinsic; and following Hume, Humean-based accounts similarly ground causality in external facts, not in intrinsic characteristics of the causal relata.

Regularity theories after Hume

Hume was considered as the denier of necessary connections in nature. This denial hand in hand with the extrinsic foundation of causal relations may be seen as the hallmarks of modern Humeanism. Empiricist philosophers after Hume followed him and abandoned any mysterious and metaphysical concept of causation. They also thought that the main problem was the traditional idea that causation implies the existence of necessary connections in nature. The project initiated by Hume and followed by other Humeans such John Stuart Mill or by Carnap and his fellow logical empiricists was to further develop these ideas.

They took over his denial of necessity in nature and tried to offer a reductive account of causation – where the truth conditions of causal relations are specified in non-causal terms. They equated causation with de facto invariable succession or actual regularity. As Carnap writes, "[a] statement about a causal relation . . . describes an observed regularity of nature, nothing more" (Carnap, 1974 p. 201). Regularity theorists claim that to the extent to which we can have an account of causation as it is in the objects, causation can only be invariable succession.

But does that mean that causation is equated with actual regularities? Do we then have to reform the concept of causation so that its new meaning is fully given by the Regularity theories with no reference to necessity?

Even Hume found this task impossible, that's why he considered causation as a "natural relation". The necessity part of causation is an integral part of our thinking and we cannot get away with it easily. The reductionist project of regularity theories that philosophers after Hume pursued is a causal realist account, it tries to capture causation as it is in the world. Hume's original theory is not a causal realist theory. His point about causation as a natural relation in the individual's mind is not necessarily part of the overall project of Regularity Theories.

Main Objections

A rather important objection to Humeanism has been that regularity is not sufficient for causation. There are many regularities in nature and not all of them are, intuitively, causal. So Humeans have tried to characterize the kind of regularity that can underpin causal relations by tying causation to laws of nature. The Humeans have to draw a distinction between the regularities that constitute the laws of nature and those that are merely accidental. But in virtue of what could we possibly draw this distinction, if we want to be faithful to the Humean empiricist epistemology?

Even Hume, in his later writings was probably not completely satisfied with his regularity account and tied together two aspects of causation, regularity of succession and counterfactual dependency. He wrote:

"*we may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second, Or, in other words, where, if the first object had not been, the second never had existed."* (Hume, 1748)

The second facet of the definition constitutes the basis of another major theory of causality, the counterfactual account. If regularity of succession is not sufficient for causation, how could counterfactual expressions of the type "if the first object had not been, the second never had existed" illuminate simple commonplace expressions like "A caused B"?

The counterfactual approach to causation is part of a broader recent development of Humean-based accounts to causation, where causes are re-defined as difference-making factors or relations of dependence. The major difference-making accounts of causation is the probabilistic approach – the claim that causes raise the chances of their effects – that is, the probability that a certain event happens is higher if we take into account its cause, than if we don't. And secondly, the aforementioned counterfactual approach to causation, which claims "if the cause hadn't been, the effect wouldn't have been either". We will now cast light to these Humean-based developments of causation and try to evaluate their contributions to the whole debate on causation.

Probabilistic Causation

Probabilistic Causation designates a group of theories that characterizes the relationship between cause and effect using the tools of probability theory. The central idea behind these theories is that causes change the probabilities of their effects. This attempt was motivated by several ideas and expectations common to the Regularity accounts.

Given the Humean dictum that all knowledge originates with human experience and the assumption that human experience is encoded in the form of a probability function, it is natural to expect that causal knowledge be reducible to a set of relationships in some probability distribution that is defined over the variables of interest. Probabilistic causality promises a solution to the centuries-old puzzle of causal discovery – that is, how humans discover causal relationships from bare empirical observations, free of any causal preconceptions. Also, in contrast to deterministic accounts of causation, physical states and physical laws need not be specified in minute detail according to probabilistic causality, because they can instead be summarized in the form of probabilistic relationships. Probabilistic causality is, thus, also equipped to deal with the modern (i.e., quantum-theoretical) conception of uncertainty, according to which determinism is merely an epistemic fiction and nondeterminism is the fundamental feature of physical reality.

What has also been a motive for supporters of probabilistic approaches to causation is a number of problems that regularity theories of causation have faced.

The first difficulty that regularity theory faces is that most causes are not invariably followed by their effects. For example, smoking is a cause of lung cancer, even though some smokers do not develop lung cancer. Regularity of succession is not always a necessary condition for causality. The central idea behind probabilistic theories of causation is that causes change the probability of their effects; an effect may still occur in the absence of a cause or fail to occur in its presence.

Secondly, a condition that is invariably followed by some outcome may nonetheless be irrelevant to that outcome. So, regularity of succession is not even sufficient for causation. Probabilistic theories of causation capture this problem by requiring that a cause make a difference for the probability of its effect.

Thirdly, if C causes E, then, typically, E will not also cause C. One way of enforcing the asymmetry of causation is to stipulate that causes precede their effects in time (as regularity theories do). But it would be nice if a theory of causation could provide some explanation of the directionality of causation, rather than merely stipulate it. Proponents of probabilistic theories of causation have attempted to use the resources of probability theory to articulate a substantive account of the asymmetry of causation (a cause C increases the probability of its effect E, but not the other way around).

Last, suppose that a cause is regularly followed by two effects. These effects are also characterized by their regularity of succession. In fact, however, the regularity relating these two events is spurious, since their regularity of succession is the effect of their common cause, but they are otherwise not causally related, they are independent. The ability to handle such spurious correlations is probably the greatest source of attraction for probabilistic theories of causation.

Reichenbach, one of the most prominent supporters of probabilistic causation, introduced the terminology of screening off to resolve this problem of spurious correlations. If C and E are both caused by some third factor, A, then it may be that C raises the probability of E, even though C does not cause E. In statistics, a confounder is a variable that influences both the dependent variable and independent variable, causing a spurious association. Reichenbach recognized that there were two kinds of causal structure in which C will typically screen A off from E (see (Reichenbach, 1956)).

The first occurs when A causes C, which in turn causes E, and there is no other route or process by which A effects E. In this case, Reichenbach said that C is causally between A and E. We might say that C is an intermediate cause between A and E, or that C is a proximate cause of E and A a distal cause of E. The second type of case that produces screening off occurs when C is a common cause of A and E. That is, if A and E are spuriously correlated, then A will be screened off from E by a common cause (The Common Cause Principle).

Objections

The original hope of Reichenbach was to provide a reduction of causation to probabilities. According to Reichenbach's official definition of probability, a probability is a limiting frequency of the occurrence of a member of one class given the occurrence of a member of the other. Probabilistic causation has given rise to 'causal modelling', which describes the new interdisciplinary field devoted to the study of methods of causal inference through probabilities. The previous discussion on the theory of probabilistic causality has conveyed some of the complexity of the problem of inferring causal relationships from probabilistic correlations.

However, what is evident from the causal modelling framework, which has been based on the foundations of probabilistic causality and describes a number of techniques that have been developed for representing systems of causal relationships, and for inferring causal relationships from probabilities, is that if we have a set of variables V and a probability distribution P on V, the probability distribution P is not sufficient to pick out a unique causal graph G on V. The observation of a probability distribution does not suffice to infer a causal relation.

Nonetheless, the theory imposes probabilistic constraints upon possible causal relations in the sense that a given set of probability relations will be incompatible with at least some systems of causal relations. The criterion for choosing an appropriate set of covariates for adjustment cannot be based on probabilistic relationships alone; it must rely on other causal information. A specification of some factors has to be held fixed and this specification appeals to causal relations, so the theory no longer offers a reductive analysis of causation. There seem to be principles we use that connect probabilistic relations to causal relations.

In conclusion, the same problem that regularity theories face in distinguishing between accidental regularities and law-like regularities is carried on by probabilistic causation. It is a common problem for general causation, since these theories are grounded on a plurality of data. General

causal claims, such as "smoking causes lung cancer" typically do not refer to particular individuals, places, or times, but only to event-types or properties. On the other hand, singular causal claims, such as "My dad's heavy smoking at his 20s caused him to develop lung cancer", typically do make reference to particular individuals, places, and times. The notion of singular causation has not reached an adequate state of conceptualization or formalization in the regularity or probabilistic account of causation and has created a tension between singular and general causes. So, probabilistic causality cannot be regarded as a program for extracting causal relations from probabilistic information; rather, it should be viewed as a program for validating whether a proposed set of causal relationships is consistent with the available probabilistic information.

Counterfactual Accounts

For the reasons stated above the probabilistic approach seems to have lost favour with a large number of philosophers. The tide now seems to have turned in favour of the counterfactual conception of causation whose initial version was found as a remark in Hume's regularity definition, but whose modern version is due to the seminal works of John L. Mackie (Mackie, 1974) and David Lewis (Lewis, 1973). This approach has recently attracted a lot of attention and there appear to be numerous ongoing research projects attempting to develop a satisfactory counterfactual analysis of causality. The core idea of the counterfactual conception is that C is a cause of E just in case E is counterfactually dependent on C – that is, if C had not occurred, E would not have occurred. The cause is rendered counterfactually necessary for the effect. Counterfactual considerations can successfully deal with many of the objections to the regularity or probabilistic theory and can distinguish causally relevant factors from mere correlations.

Counterfactual accounts although Humean-based, since they do not offer an intrinsic analysis of causation and regularity still plays a central role as we will see, also tried to capture the meaning of singular causal statements and to resolve the tension between singular and general causation that regularity (and probabislistic) causation create.

Mackie's counterfactual account

Mackie's counterfactual account was developed as an attempt to fill the gaps that regularity theories left. He tries to capture the meaning of singular causal statements in addition to general causation and to offer a realist account of causation; "as it is in the objects". In this section we will explore his attempted unification of a regularity-based theory with singular causation. Later on, we will also consider his ideas on the intrinsicness of causal relations.

What is the meaning of a singular causal statement? When, for instance, we say that one particular event was caused by another one, what do we exactly assert? Mackie argues that a causal statement of the form "c caused e" should be understood as follows:

c was necessary in the circumstances for e, where c and e are distinct event-tokens.

And he adds, as an explanation of this, the following counterfactual assertion:

if c hadn't happened, then e wouldn't have happened (Mackie, 1974 p. 31).

Humeanism according to Mackie fails to capture the meaning of singular causal statement and through counterfactuals such as the above he tries to capture the concept of singular causation. The missing meaning of a singular causal statement can be fixed by the appropriate counterfactual conditionals (Mackie, 1974 p. 270).For Mackie general causal statements are "quantified variants of the corresponding singular ones" (Mackie, 1974 p. 80).

Counterfactual statements are not, according to Mackie, true or false as such. They do not describe "a fully objective reality". So, under what circumstances is a counterfactual conditional true or false?

Mackie views counterfactual statements as potentially true or false assertions, whose truth-conditions depend on the inductive evidence that supports them (Mackie, 1974 pp. 229-230). As Mackie notes, the evidence plays a double role. It first establishes inductively a generalization and then, "it continues to operate separately in making it reasonable to assert the counterfactual conditionals which look like an extension of the law into merely possible worlds" (Mackie, 1974 p. 203). So the truth of counterfactual statements depends on inductively established generalizations. If, in the actual world, there is strong evidence for the general proposition "All Fs are Gs", then we are justified to extend this generalization beyond the observed facts "to merely possible ones" (Mackie, 1974 p. 55). We believe – due the evidence we have – that "if x had been an F it would have

been a G". It is important to point out that Mackie is no realist about possible worlds as Lewis is. Mackie does not think that possible worlds are real, but he still thinks that it is convenient to talk in terms of possible worlds in our assessment of counterfactuals.

Objections

A serious objection to Mackie's account is the fact that his supposed singular causal statements have actually no adequate truth conditions. Since Mackie's account of the meaning of singular causal statements depends on counterfactuals, and since counterfactuals are not strictly true or false, but depend on generalizations, it follows that singular causal statements cannot be objectively true or false (Mackie, 1974 p. 54). One of the important claims of those who defend singular causation is that regularity theories fail to offer adequate truth conditions to singular causal statements. In this respect, Mackie's theory fails as well.

Secondly, although Mackie has aimed to identify an intrinsic, objective feature of a causal sequence of events that makes the sequence causal, he has also failed to do so. Whether a sequence of events will be considered as causal depends on the evidence which supports the relevant counterfactual conditional, namely, on an extrinsic feature. The distinction between causal sequences of events and non-causal ones is epistemic; it depends on whether the evidence is strong enough to support through a generalization the relevant counterfactual. As he says: "the holding of a counterfactual conditional is not a fully objective matter" (Mackie, 1974 p. 55).

This is the reason why he developed his theory into including a mechanistic approach to further search for the intrinsic feature of causation. We will review on this development of Mackie's account in a later chapter on mechanistic approaches.

Lewis's theory

David Lewis is the best-known advocate of the counterfactual approach to causation. In his book "Counterfactuals" (Lewis, 1973) he offered a counterfactual theory of causation under the

assumption of determinism. Unlike Mackie, Lewis puts forward an objectivist theory of counterfactuals, based on semantics of possible worlds. His theory is moreover a singularist account of causation.

Causation according to Lewis is a relation of counterfactual dependence between events. He analyzes this counterfactual dependency in terms of similarity relations among possible worlds. Lewis uses regularities, but only in order to ground his argument of similarity relations among possible worlds. So, his account of causation is not – strictly speaking – a regularity theory.

Lewis' counterfactual theory tries to explain what we should envision as changed and what should be held fixed when we evaluate a counterfactual the antecedent of which is not true of the actual world. Within Lewis' framework, this is the issue of which worlds, in which the antecedent of the counterfactual holds, are "closest" or "most similar" to the actual world. Lewis' answer to this question invokes a "similarity" ordering that ranks the importance of various respects of resemblance between worlds in assessing overall similarity (Lewis, 1986 p. 163). The actual world, according to Lewis, is one among the many possible worlds, which are no less real than the actual. Actual and possible worlds differ from each other in some facts, or in some laws. The actual world is just the one which is comprised of the totality of facts in the actual world. For Lewis it is therefore possible to rank worlds according to how similar they are in respect to facts and laws. When the antecedent of a counterfactual is not true of the actual world, Lewis' similarity metric commonly leads us to think of that antecedent as made true by a miracle. Small miracles allow for small violations of the laws of nature and the distribution of particular facts in comparison to the facts in the actual world, whereas big miracles allow for big violations of facts and laws and are thus indicative of a greater difference compared to the actual world.

Lewis' idea is that the appropriate counterfactuals for analyzing causation are often counterfactuals the antecedents of which are made true by "small miracles". As Lewis put it: "a counterfactual . . . is true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent" (Lewis, 1986 p. 164). His intention is to ground the truth-conditions of counterfactuals in an objective way. "Small miracles" is a notion that functions so as to provide the antecedent event C with the kind of "independent causal history" that allows us to distinguish the effects of C on E from the effects of other "confounding" variables on E. Lewis' account, with its somewhat ad hoc looking similarity

ordering, works as well as it does because his similarity ordering picks out roughly those relationships that are stable under interventions and distinguishes them from mere correlations. As we will see in the next section, according to the interventionist account, it is just these relationships that are causal. But how do we get causation out of all this?

In order to ground his causal theory on counterfactuals, Lewis outlines some steps (Psillos, 2002 pp. 93-94) The first step is to define a notion of counterfactual dependence between families of propositions, each of which describes the occurrence of events. To every event e, there is a proposition O(e) and a proposition not-O(e), which state whether e occurs or not. Secondly and most importantly, Lewis has to define causal dependence in terms of counterfactual dependence. So e causally depends on c iff the family of propositions O(e) and not-O(e) counterfactually depends on the family O(c) and not-O(c). And finally, he links causation with causal dependence: "Causal dependence among actual events implies causation" (Lewis, 1986 p. 167). If two events c and e are actual, and e is counterfactually dependent on c, then c is the cause of e. The conditional "If c had not occurred, e would not have occurred" entails the causal statement "c caused e". According to Lewis, causal dependence between actual events is sufficient for causation.

A success of Lewis' account of causation is that it can deal with cases of pre-emption. A pre-empted event Y is an event that would have led to a certain event E, but it is such that its occurrence is blocked (or pre-empted) by the occurrence of another event C, which nonetheless causes E. Given the definition of causation in terms of causal chains (Lewis, 1986), Lewis is able to distinguish preempting actual causes C from preempted potential causes Y. However, there is number of objections to Lewis' account.

Objections

A serious cause of concern with Lewis's theory stems from his account of the truth-conditions of counterfactuals in terms of the semantics of possible worlds. For most of us, there is only one world; namely, the actual and even if possible worlds exist, it is rather speculative to base our account of causation on them. So Lewis's conceptualization of possible worlds is a risky metaphysical commitment, which cannot be easily embraced. Added to that, what criteria should we

employ in order to rank possible worlds according to their similarity to the actual one? Part of the ranking will be left on our subjective judgements and the intended objectivity of the counterfactuals' truth-conditions will start to shake. The ranking criteria, which Lewis (Lewis, 1973) introduces, are not clear enough and they complicate the theory unnecessarily.

There is another very central objection to Lewis' theory, which actually applies to all counterfactual approaches and we will extendedly discuss it in later chapters. It comes from the possibility of causal overdetermination. Overdetermination means that an effect has more than one sufficient cause, but none of them is necessary for the effect, since if one does not occur, the other(s) would ensure the occurrence of the effect. This shows that either there is causation without counterfactual dependency or that there is a plurality of counterfactual dependencies in these cases. How can a counterfactual account deal with them?

Lewis (Lewis, 1986 p. 194) replies that cases of overdetermination are not a threat to his counterfactual account, since we cannot apply causal terminology to such cases. Does that mean that none of the events on which the effect counterfactually depends causes it? It is a problem for his own theory that, in cases of overdetermination, there is no answer to the question of whether each of the causally sufficient factors (which we cannot call causes) *caused* the effect (Horwich, 1987). Cases of overdetermination raise difficulties for his analysis; they constitute prima-facie evidence that counterfactual dependency is too narrow to capture causal dependency.

A further observation, made by Kim (Kim, 1973), is the fact that there are also cases of counterfactual dependency without causation. A strong example is Kim's example of determination without causation: "When my sister gave birth to her first child, I became an uncle. My becoming an uncle was determined by, was dependent on, the birth of the child, but was not a causal effect of it" (Kim, 1973 p. 571). Another example indicates cases where one event is constituent of another and therefore counterfactually depended on it, without them being causally related though (Kim, 1973 p. 571). If these cases are plausible, counterfactual dependency can be considered too broad to describe causal dependency.

Interventionism

Recently, there have been some developments of the counterfactual approach, in which causation is linked with the notion of manipulation or intervention (Menzies, et al., 1993) (Woodward, 2000) (Hausman, et al., 1999). The manipulability/interventionist theory can be regarded as a new variant of the counterfactual view of causation and it also belongs to the causation-as-dependence family. The theory is quite sophisticated and flexible, and can deal with many well-known problems of the more traditional counterfactual theory.

The underlying idea is nevertheless quite simple. It lies upon a basic epistemological inquiry: how do people ever acquire causal information from the environment and, more specifically, how children extract causal information from experience. Hume's causal theory has a similar empirical point of departure. However, he ignored one important fact: that the child never operates in a closed, isolated environment. External conditions govern the operation of every learning environment, and these conditions often have the potential to confound cause and effect in unexpected ways. But fortunately, aside from passive observations, a child possesses another valuable source of causal information that is not available to the ordinary statistician: manipulation. Via manipulation the putative causal event is left to the sole influence of a known mechanism, thus overruling the influence of other uncontrolled factors and conditions. The beauty of independent manipulation is, of course, that these other factors can be kept constant without being identified (Cheng, 1992). The independence is accomplished by subjecting the object of interest to the whims of one's volition in order to ensure that the manipulation is not influenced by any environmental factor likely to produce the putative effect. Manipulation serves to filter effects produced by the child's actions from those produced by uncontrolled environmental factors. It replaces the statistical notion of randomized experimentation.

Manipulability/Interventionist theories are based exactly on this idea; that causal relations are potentially exploitable for purposes of manipulation and control. I will here focus on the interventionist version of these developments developed especially by James (Woodward, 2000) (Woodward, 2003) since it can be also applied where human agency is absent and therefore is a more objective account of causation. It is the theory of causation that is now becoming prevalent.

Very roughly, according to interventionist theories, if C is genuinely a cause of E, then if I can intervene on C in the right way, this should be a way of changing E. An intervention I on C that is followed by changes in E is a sufficient condition for C to cause E (Hausman, et al., 1999 p. 537).A relationship among the variables C and E is said to be causal if, after an appropriate intervention on C, the relationship between C and E wouldn't change, but the value of E would change.

The value of the variables change, the relation ought to remain unchanged. This is why Woodward links the notion of intervention with the notion of invariance (Psillos, 2004 p. 295). A certain relation (or a generalization) is invariant, Woodward says, "if it would continue to hold – would remain stable or unchanged – as various other conditions change" (Woodward, 2000 p. 205). The notion of invariance refers to the causal relation, which instantiates a generalization and is not a feature of a singular sequence. The causal relation should remain stable under a set of actual and hypothetical interventions. In that sense the interventionist theory is a counterfactual approach, since what matters is what would happen to a relationship if interventions were to be carried out.

Causes are defined as "levers that can be used to manipulate their effects" (Hausman, et al., 1999 p. 533) and an effect is defined in terms of the difference made to its value by a change in the value of the cause. Focusing on changes in this way allows us to isolate the contribution made to E by C alone from the contribution made to E by its other causes. But in order to ground this isolation of the C-E relation, it is required that an intervention on C leaves intact all other mechanisms besides the mechanism that previously determined the value of C.

According to Woodward an intervention I on C counts as an intervention if the change of the value of C is entirely due to the intervention I (I is the only cause of C; all other causes are erased); and the intervention I changes the value of E, only through changing the value of C. These requirements make sure that the change of C does not have causes other than the intervention I, and that the change of E does not have causes other than the change of C (and its possible effects). In this way it is ensured that E-changes are exclusively due to C-changes, which, in turn, are exclusively due to the intervention I (Psillos, 2004 pp. 294-295).

Whatever the true theory of causation looks like exactly, we have noted in particular some general philosophical points that emerge from the critical discussion so far. First, one should clearly distinguish between causation and correlation and second, genuine causes are difference-makers for

their effects and thus should not to be confused with mere sufficient conditions. Both of these points are successfully grasped by the interventionist theory.

Causation-as-dependence is arguably not an absolute relation; causal judgments depend on contexts and contrasts. The interventionist approach relates variables, which can take different values. Different choices of possible alternative values result in different contrast classes. The default contrast is that the presence rather than absence of the property/event at issue is caused by the presence of another property/event rather than absence of it (Woodward, 2003 pp. 67-68). In this way, the interventionist theory embodies the important idea that causal claims are essentially contrastive.

The interventionist theory also seems to erase the artificial tension that has been created between general causation and singular causation. This distinction plays an important role in understanding the nature of causal explanations, but the two types of causation differ merely in the level of scenario-specific information that is brought to bear on a problem. By accommodating counterfactual conditionals to calculate the effect of an intervention, singular and generic causes no longer stand in need of separate analyses. They are just different ways of contextualizing propositions.

Objections

One limitation of the interventionist account concerns the scope of the requirement, that an intervention on C leave intact all other mechanisms besides the mechanism that previously determined the value of C. Suppose that variables C and E are parts of a greater system, which is governed by other causal laws in addition to the causal law that connects C and E. There are cases where an intervention I on C would not only change E, but would also independently change some other causal law of the system, which in turn has an influence on E. In this case, we cannot clearly conclude as to which is the genuine cause of E. (Psillos, 2002 pp. 104-105)

Hausman and Woodward (Hausman, et al., 1999 p. 542) introduced the concept of modularity in order to ensure that an intervention on the variables of a causal law do not have an effect on the other possible causal laws of the system. Hausman and Woodward explain modularity by claiming that "if two mechanisms are genuinely distinct it ought to be possible (in principle) to

interfere with one without changing the other". Yet as Nancy Cartwright (Cartwright, 2000) has persuasively argued, modularity is a very strong requirement, which breaks down often. So, one first limitation of the interventionist account concerns the limitations of the modularity requirement.

A second general point about the interventionist account is that it provided just an inference tool and not a definition of causation. It can be argued that Woodward fuses his epistemology with the metaphysics of causation (Psillos, 2002 p. 100). It is questionable if the causal relation between X and Y merely consists in this inference tool of interventions. The invariance observed under the interventions on the cause could be just a symptom of the causal relation, which is an intrinsic feature of the causal relata. Woodward seems to agree on this objection: "what matters for whether X causes . . . Y is the 'intrinsic' character of the X-Y relationship but the attractiveness of an intervention is precisely that it provides an extrinsic way of picking out or specifying this intrinsic feature" (Woodward, 2000 p. 204). But he goes further by noting that "for Y to change under an appropriate intervention on X just is what it is for X to cause Y" (Woodward, 2000 pp. 205-206). Invariance under interventions is all that there is in causation, according to Woodward.

Added to this, the notion of an intervention cannot be used in order to provide a reduction of causal claims to non-causal claims, since an intervention is itself a causal term (it causes changes on X). So, even if we accept this approach, we must be aware that such a theory is at least non-reductionist.

What about the truth-conditions of causal statements in the interventionist theory? Woodward aims at a simultaneously pragmatist and realist account (Psillos, 2004 p. 297). On the one hand, the truth-conditions of counterfactual statements are not to be specified by means of an abstract metaphysical theory, such as Lewis's theory; on the other hand, he does not want to say that the truth-conditions of causal statements are fully specified by actual and hypothetical experiments.

He insists that counterfactual causal statements have their truth-conditions independently of their evidence-conditions. But only those counterfactuals related to interventions can tell us which relations are causal. In his (Woodward, 2003 p. 122) he stresses that "the appropriate counterfactuals for elucidating causal claims are […] those that have to do with the outcomes of hypothetical interventions. [. . .]". So, it is in virtue of the interventions, that the truth-conditions of counterfactuals are tested. To sum up, counterfactuals have a truth value independently of the

evidence we have for it, but we conduct the experiments in order to discover whether it is true or false (Woodward, 2003 p. 123).

As we can observe, his intention to offer a simultaneously pragmatist and realist account faces a difficulty. In order to ground the truth-value of counterfactual conditionals, he either has to take a pragmatist approach of evidence-based counterfactuals or take a realist stance, were counterfactuals are true in virtue of an intrinsic feature or a mechanism, which the evidence discovers. We will now turn to the mechanistic accounts, which try to discover or uncover the intrinsicness of causal relations.

Causation as production: Mechanistic Approaches

Causation-as-production, as we already explained, is the second big family of causal theories. Up to now, we have kept ourselves busy with the causation-as-dependence family and its problems. According to dependence theories, to say that c causes e is to say that e suitably depends on c. According to causation-as-production theories, c causes e means that something in the cause produces or generates the effect. As dependence can take many forms, such as nomological, probabilistic or counterfactual dependence, similarly production can be explained in different ways. The most popular view is the mechanistic approach, which states that there is a mechanism that links the cause and the effect; two events are causally related if and only if there is a mechanism that connects them.

Humean-based (or humean-inspired) theories intentionally overlook an important aspect of causation; namely its intrinsic, necessary aspect. Since this necessary tie between cause and effect is not observable, Humeans are not interested in grounding their causal theories on it. However, some philosophers attempted to show that there is a tie between cause and effect, which consists of a mechanism. This idea seems very attractive considering the problems that humean-based approaches face when it comes to the metaphysics of causation (i.e. the epistemic fallacy problem). But can it be made to work?

Mackie's mechanistic approach

Mackie's counterfactual approach was an attempt to ground the meaning of singular causation, but as we have seen, he has failed to do so. The main objection to his counterfactual approach is that he could not give an objectivist account of the truth-values of singular causal claims. Whether a sequence of events could be considered causal depends on whether there is evidence to support the relevant counterfactual conditional. He therefore develops a mechanistic approach in order to explain causation "as it is in the objects" and in that way capture the meaning of singular causal statements. Mackie's mechanistic theory was the first contemporary approach in this direction and he was the one and only philosopher to offer a counterfactual and a mechanistic approach to causation in parallel. Since then these theories have split parts (Psillos, 2002 p. 39).

Mackie's main point is that there is an intrinsic feature of each singular causal sequence, which distinguishes causal from non-causal sequences (Mackie, 1974 p. 20). Hume might be wrong that there are no evident causal powers. Mackie thought so and attempted to give an account of the intrinsic feature existing in each singular causal sequence.

This feature or mechanism, he argues, consists in the qualitative or structural continuity, or persistence, exhibited by certain processes, which can be held causal (Mackie, 1974 pp. 218-220). The feature (or structure) that persists differs among causal processes. For instance, what persists can be "the total energy" of a system, or the "number of particles", or "the mass and energy" of a system (Mackie, 1974 pp. 217-218). This feature which persists in a certain process is what connects together the several stages of this process and renders it causal.

Mackie does not go in detail as to what exactly this structural continuity of a process is. He is just confident that in each sequence of events there will be microscopic descriptions of these events in terms of their structure (Psillos, 2002 p. 110), which will show that there is more continuity and persistence in the sequence than observable. However, Mackie does not explain how a process of change can be reduced to a process of persistence. Persistence leaves out one important aspect of causation; namely, causal interaction. When two processes interact there should be something more than just the structural persistence.

Salmon's mark transmission

Wesley Salmon has also tried to offer a systematic account of the mechanism that links cause and effect. His early attempt to analyse physical causation was in terms of what he called 'mark transmission'. Later on he adapted Dowe's notion of transference of conserved quantities, which we will explore shortly after Salmon's mechanistic approach. Salmon stresses that there is a local tie between a particular cause and a particular effect and he also tried to analyse systematically certain basic causal concepts in terms of non-causal ones in order to offer a reductive theory of causation.

Salmon, following Mackie, takes causal processes to be the fundamental element of the mechanistic approach to causation. Processes, instead of events, "are the mechanisms that propagate structure and transmit causal influence in this dynamic and changing world. . . . they provide the ties

among the various spatiotemporal parts of our universe" (Salmon, 1997 p. 66). The continuity of the process accounts for, ultimately, the direct link between cause and effect (Salmon, 1984 pp. 156-157).

As Salmon describes it, a process is something that transpires over a period of time. A material object in motion is a process; so is the propagation of a sound wave or a pulse of light. A moving shadow is a process, so is the image of a golf ball on a TV screen. A material object at rest is a process. Roughly, a process is something that is represented in a spacetime diagram as a line; an event, in contrast, is represented by a point (Salmon, 2010 p. 7).

However, not all processes are causal. Salmon, borrowing an idea of Reichenbach, characterized as causal only those processes that are capable of transmitting a mark (Salmon, 1984 p. 142). What does it mean to transmit a mark? To mark a process is to interact with it and "stamp" it. Mark transmission, though, means that not only should the process be marked, but that the mark gets transmitted as well. For Salmon without the transmission of the mark, the process is not causal.

A mark, Salmon adds, is a modification of the structure of a process. Every process exhibits "a certain structure", but only a causal one is capable of transmitting its own structure, thus modifying the overall structure in a single local interaction (Psillos, 2002 p. 112). Salmon writes: "if a process – a causal process – is transmitting its own structure, then it will be capable of transmitting certain modifications in the structure" (Salmon, 1984 p. 144). It seems that a causal process is an interaction, where a mark is being transmitted. But isn't the concept of interaction itself causal?

Since Salmon's intentions were to offer a reductivist theory of causation, he tried to define the concept of causal interaction in non-causal terms. He defines it in terms of the geometric (i.e. non-causal) concept of intersection of two processes. The concept of causal interaction "*is based on the geometrical notion of an intersection of worldlines of processes (causal or pseudo) in four dimensional spacetime. We can say that an intersection of two processes is a causal interaction if and only if each process undergoes a change in the locus of intersection which persists beyond the intersection. (The word "causal" is redundant; all interactions are causal. Those intersections that fail to qualify as causal are simply intersections.)*" (Salmon, 2010 pp. 7-8). After all, it might seem that Salmon indeed managed to offer an analysis of causation in non-causal terms, as he has wished for.

Objections

Salmon's definition of interaction in terms of the non-causal, geometric concept of intersection reminds us of Hume's spatiotemporal contiguity and constant conjunction account. Salmon himself points out that the "ability to transmit a mark can be viewed as a particularly important species of constant conjunction – the sort of thing that Hume recognised as observable and admissible" (Salmon, 1984 p. 147).

However, as Psillos points out, there is a major difference. Salmon talks about the *ability* to transmit a mark. He wants in that way to guarantee that a process can be regarded as causal, even if it is not actually marked (Psillos, 2002 p. 115). Unlike Hume's theory, Salmon's theory is based on a dispositional feature of a process. This point is quite important, since non-actual processes can only be analysed by hypothetical counterfactual statements, which means that Salmon's account depends on counterfactuals. We will shortly return to this point.

There is another difficulty that Salmon's account faces concerning his definition of mark transmission. He argued that to mark a process means to interact causally with it so that its structure is modified. And then he defined causal interaction in terms of structure modification by means of a process, which is also marked (Psillos, 2002 p. 117) So, the definitions of causal interaction and of marking seem to be mutually dependent resulting in circularity (Dowe, 2000 p. 72). This circularity problem lead Salmon to adopt Phil Dowe's conserved quantity approach.

Dowe's conserved quantity approach

Phil Dowe also aims to offer an empirical physical theory of causation as it is in the actual world. He has modified Salmon's mechanistic theory and instead of a mark transmission process, his theory relies on the concept of a conserved quantity (Dowe, 2000 p. 89).

A conserved quantity, as Dowe explains, is "any quantity that is governed by a conservation law" (Dowe, 2000 p. 91). Examples of such quantities can be found in our best scientific theories; mass, energy, linear momentum and charge are some of them. Whether a process is causal will depend on whether it possesses a conserved quantity. Dowe insisted that what renders a process causal lies in the "possession" of a conserved quantity, whereas Salmon's endorsement of Dowe's

theory argues for a "transmission" of a conserved quantity. Is Dowe's "possession of a conserved quantity" sufficient for causation?

For Dowe possession of a conserved quantity is identical to the instantiation of it (Dowe, 2000 p. 92). So to say that an object possesses a certain amount of mass or an electrical charge means that these properties are instantiated by this object. Typical cases of non-causal processes (such as shadows) possess no conserved quantities, so will be considered as pseudo-processes by Dowe's theory. In particular, Dowe does not require either that the conserved quantity be kept constant (as Mackie thought) or be transmitted (as Salmon claimed). In this way, Dowe's mechanistic account manages to escape the obstacles that Mackie's and Salmon's theories face.

Objections

Dowe's theory, as well as Salmon's later modified theory, intends to analyse causation without an appeal to counterfactuals and if we look at their definition of a causal process, it seems that they managed to do so.

But let's consider again the case of non-actualized, potential causal processes. Let's take the example of a chemical reaction, which has not yet taken place, since the chemical elements have not been actually brought into contact, in order to causally interact or exchange their conserved quantities (Psillos, 2002 p. 126). Do they share the same causal status with actual processes, which however do not possess any conserved quantities (such as shadows)? Non-actualized causal processes do not possess conserved quantities, so according to Dowe's definition are not considered as causal. But in comparison with other non-causal processes, they would possess conserved quantities, if they were actualized. Do both cases equally share the status of non-causal processes? The only way to compare a non-actualized, but potentially causal process with a de facto non-causal process is by making use of counterfactual conditionals, by claiming that "if they were instantiated, they would then possess a conserved quantity".

A second general point made by Hitchcock, is that Dowe's and Salmon's theories have serious difficulties in identifying the causally relevant quantities. As he points out (Hitchcock, 1995), often in causal interactions, several conserved quantities are exchanged or possessed. Consider, for

example, the case of a pool cue which strikes a cue ball (Raatikainen, 2018 p. 31). Both momentum and a blue dot of chalk are transmitted. In the former case, momentum is exchanged. In the latter, matter is exchanged. Yet only the first is relevant to the trajectory of the cue ball. It is unclear how to determine which exchanges are relevant. A natural response is to rely, again, on counterfactuals and find the difference-making factor. If we had removed the dot, or had changed its colour, the trajectory of the ball would have stayed the same. Thus, it seems that an ultimate avoidance of counterfactuals is not possible. But this solution is not open for Salmon and Dowe, who are trying to avoid them.

Glennan's Mechanisms

Both, Salmon and Dowe, characterize interactions in terms of the exchange of conserved quantities. They aim at a mechanistic theory of physical causation; their account is too narrow to describe cases of causation among higher-level entities though.

Mechanisms can also be more complex systems, such as the mechanism of reproduction in biology or the function of an organ in our body or a constructed machine or tool, such as the thermometer. These more complex mechanisms bring about a certain activity or are responsible for a certain behavior. Stuart Glennan's mechanistic approach is attractive, since it aims at a more adequate description of higher-level mechanistic causation.

According to Glennan (Glennan, 2002 p. 344):

(M) *A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.*

A mechanism consists of several parts, but it is the way these parts are organized, which determines how the parts interact with each other to produce an activity or a behavior. The properties of the parts, as well as their organization, should be stable. Briefly, the system as a whole should have stable dispositions, which produce the behavior of the mechanism. A mechanism is more than the sum of its parts, since each part contributes to the overall behavior of the mechanism more than it would have achieved if it acted on its own (Psillos, 2004 p. 310).

Glennan accepts the supervenience of the interactions involved in higher-level mechanisms on fundamental physical interactions, but he thinks it is not helpful to describe these mechanisms in terms of them. We would surely miss something if we tried to explain a complex man-made machine, say a smart phone, in terms of the exchange of conserved quantities. We would "lose the fact that higher-level interactions form higher-level kinds" (Psillos, 2004 p. 308). On the debatable topic of the autonomy of higher-level sciences and laws, we will reflect in the final section of this thesis in an attempt to explain mental phenomena. For now, we can agree that Glennan's mechanistic view is broader than this of Salmon and Dowe and explains mechanisms at levels higher than fundamental physics.

A central feature of Glennan's views is a distinction between the fundamental laws of physics and mechanically explicable laws, which are laws grounded on a mechanism, or as Glennan says, which are "explained by the behavior of some mechanism" (Glennan, 1996 p. 62). Special sciences are characterized by mechanically explicable laws and not by the fundamental laws of physics (Glennan, 1996 p. 50).

Moreover, Glennan argues that his mechanistic approach offers a solution to the much-discussed problem of the truth-conditions of counterfactuals. The reason why some counterfactual causal claims are stable, is because they are supported by mechanically explicable laws (laws that are based on mechanisms) (Glennan, 1996 p. 63). The presence of the mechanism (e.g., the human heart) explains why a certain counterfactual holds, (e.g., if the heart suddenly stops beating, life cannot be sustained, since the heart as a mechanism provides the body with oxygen-rich blood).

Objections

However, if some counterfactuals are true even though a mechanism is absent, then there is more to the link between laws and counterfactuals than Glennan's theory admits and his theory is incomplete. And we have seen that Glennan also accepts the supervenience of non-fundamental laws on the fundamental physical laws (Glennan, 2002 pp. 346-347). So on Glennan's view, non-fundamental laws are based on mechanisms and supervene on fundamental laws, which are not based on mechanisms.

So, what is the relation between the mechanisms that realize the non-fundamental laws and the more fundamental laws on which the non-fundamental laws supervene? If supervenience holds, the mechanism and its laws have no metaphysical autonomy in itself and the fundamental physical laws ought to be part of the truth-makers of the non-fundamental laws (Psillos, 2004 p. 310). Once identified, the mechanism might well have explanatory and epistemic autonomy, but the fully sufficient truth-conditions of certain counterfactuals must also include some facts about the fundamental laws, on which the behavior of the mechanism supervenes.

Mental Causation Problems and Causal Theories

We have finally come to the final section of this thesis. Up to here we have explored some of the central contemporary problems and debates of mental causation; and consequently, we encountered the major theories of causation and discussed their strengths and weaknesses. It is time, to engage with the application of the later in order to tackle the former. Shortly, we will now examine how our choice of one causal theory contrary to another can have different outcomes in our analysis of mental causation problems.

The next part will tackle some common mental causation problems, which arise from the embracement of a non-reductive, physicalist perspective and see how different theories of causation deal with them. The final part is dealing with the consequences of adopting a reductionist view and how causal theories explicate them.

Non-reductive Physicalism & Causal Theories

Let's begin with the endorsement of a physicalist outlook; we think that everything is physical or supervenes on the physical. However, we also accept that certain higher-level properties or events (like mental or biological ones) play an important causal role. For any such higher-level phenomena, there are physical conditions metaphysically sufficient for their presence; but, those higher-level properties are not identical with, or reducible to, their underlying physical conditions. This then seems to commit us to accepting systematic overdetermination as a common feature of the world since any effects caused by the higher-level phenomena would also be caused by their underlying physical bases.

This sort of problem has been claimed to arise for antireductionists in a number of domains, but it has been most intensively discussed in connection with the mental causation problem, as we already discussed in the first section of this thesis. Kim has used a version of it, known as the "exclusion problem," to pressure his fellow physicalists toward reductionism (Kim, 2005). The

exclusion argument applies to all forms of property dualism that accept physical causal closure according to Kim and goes like that.

Anti-reductionists claim that mental properties are distinct from, and irreducible to, physical properties. Now, Kim adds to this the following plausible metaphysical thesis of causal exclusion: *No event can have more than one sufficient cause occurring at any given time – unless it is a genuine case of causal overdetermination.* (Kim, 2005 p. 42).

Given these two propositions as premises, Kim intends to show that property dualism runs into trouble with mental causation. If all mental events have physical realizers or are supervenient on physical events, then all cases of mental-to-physical causation are cases of systematic overdetermination; they involve a sufficient physical cause as well as the mental cause. At this point the causal-exclusion principle applies, since the mental-to-physical causation is not a case of genuine overdetermination (the case where an event has more than one independent, sufficient causes), and either the mental cause or the physical cause must be excluded. Excluding the physical cause is not possible according to the principle of physical causal closure. The mental cause, therefore, must be let go. Any putative mental cause of a physical event is always 'excluded' by a physical cause. Given that the underlying physical base does all the work in bringing about the effect, what further causal work is left for the mental event (or the mental property of the event) to contribute (Kim, 1998 p. 53)?

The overall implication of the argument, given that we accept it, is that mental events and properties and, more generally, all events in the special sciences are epiphenomenal with respect to physical events. According to Kim we are left with a dilemma between epiphenomenalism and overdetermination, unless we accept some form of reductionism.

The Problem of Overdetermination & Causal Theories

Does it matter, for the issue of mental causation, and for the overdetermination/exclusion problem in particular, what theory of causation is presupposed? On the one hand, the exclusion argument has been presented without reference to any specific views about causation and responses to it have been also put forward as if they held quite independently of any particular theory of

causation. Kim has also presented the exclusion argument as if it stands independently of any specific theory of causation, even though in his later works he is showing increasing sympathy towards the causation-as-production view of causation (Kim, 1998 p. 45). In his recent work (Kim, 2014), Kim suggests that the exclusion argument succeeds in establishing the same conclusion, whether one understands causation as nomological sufficiency, as counterfactual dependence, or as production or generation.

On the other hand, others have suspected that the force of the exclusion problem is sensitive to how we think about causation (Raatikainen, 2018) and that all that matters is the choice of the theory of causation. It has been suggested that given the dependence view of causation, mental causation is no problem at all, whereas if the production view of causation is assumed, mental and other higher-level causation is immediately impossible even without any exclusion argument.

In order to shed light on this dispute we will now proceed with an analysis of the overdetermination/exclusion problem based on the up-to-date picture of causal theories we have presented in the previous section.

Overdetermination & Causation-as-Production

It might be that we worry about causal overdetermination, only because we think of causation as involving the transfer of some sort of matter or energy (Loewer, 2002). This conception of causation as production, however, may be false and we may have to question our worry about overdetermination and its underlying conception of causation.

Some counterfactualists believe that there is in fact no overdetermination "problem" to worry about. Barry Loewer (Loewer, 2002) pursues this line. He argues that once we reject the "production" notion of causation in favor of a counterfactual (dependence) conception, it should not worry us that our world exhibits massive overdetermination, since an event can easily counterfactually depend on multiple causes.

Kim himself admits that "Loewer is right… in saying that my thinking about causation and mental causation involves a conception of causation as 'production' or 'generation'"" (Kim, 2002 p. 675). He thinks that the counterfactual theory of causation cannot be the ultimate truth about causation, although it speaks to our intuition.

Kim argues that counterfactuals depend on laws and regularities, and that, consequently, the counterfactual approach in fact reduces to the regularity theory and shares its well-known problems: "embracing the counterfactual approach to causation will have no advantages over the regularist-nomological approach … the counterfactual approach seems to presuppose, or collapse to, the nomological conception and thereby inherit the latter's shortcomings" (Kim, 2010 pp. 233-236).

Kim invites us to consider the following counterfactual claim:

(C) *If this match had not been struck, it would not have lighted.*

According to the counterfactual theory, whether the causal claim "The striking of the match caused it to light" is true, depends on whether or not the above counterfactual conditional is true. But how do we know that the counterfactual is true? Kim argues that the truth-conditions of counterfactuals are the actual regularities in nature. It is "our knowledge of this regularity, or law, combined with knowledge of the actual circumstances in which the match was struck (e.g., it was dry, oxygen was present, etc.)" that according to Kim accounts for our judgment that (C) is true (Kim, 2010 p. 234). Kim is thus apparently claiming that our knowledge of counterfactual conditionals (*If c had not occurred, e would not have occurred)* reduces, in general, to knowing the regularity that events of type C are followed by events of type E.

This also affects the use of counterfactuals to account for mental causation. Consider the following counterfactual conditional (Kim, 2010):

(D) *If Susan had not had the sudden migraine headache, she would not have experienced frightful anxiety.*

Our commonsense knowledge that counterfactuals like (D) are often true, grounds our belief in mental causation. But what makes (D) true? If our previous observations relating to counterfactuals conditionals are correct, the truth of (D) must depend on the regularity connecting sudden attacks of migraine headaches and feelings of anxiety.

Counterfactuals regarding mental causation cases involve reference to psychological or psychophysical regularities. Kim's worry is that their significance for mental causation depends on the question whether these regularities are also causal regularities. Even an epiphenomenalist could accept a regularity of this kind and acknowledge it to be lawlike, although the epiphenomenalist will deny that the attack of migraine headache caused the sense of fearful anxiety. He would rather argue that the observed regularity arises out of the genuine causal process connecting two neural substrates on which the headache and the anxiety are epiphenomenal to (Kim, 2010 pp. 243-263).

According to Kim, the same applies even if one describes the relationship between mental states and the underlying neural states in terms of supervenience or multiple realization. The question is not whether these psychological (or psychophysical) regularities or laws exist; but rather, whether these regularities are causal regularities and whether these laws are causal laws. Horgan, for example, believes that mental events and states have the causal efficacy that they have because their neural or physical realizers have causal efficacy. A mental state, occurring on a given occasion, in virtue of being realized by a certain neural or physical state, has exactly the causal powers of that physical state.

Kim agrees up to this point with Horgan, but he further claims that the next step we are compelled to take is to reductively identify this particular mental state with its neural or physical realizer (Kim, 1998). So, he actually denies the "nonreductive" part of nonreductive physicalism. To claim that the mental state is a distinct state from its neural realizer, and yet consider the neural realizer a sufficient cause of the effect, means for Kim to accept overdetermination (Kim, 2010). And to say that the mental state has causal efficacy "via" the causal efficacy of its neural realizer carries an apparent epiphenomenal implication. What real causing is left for the mental state?

Kim argues that neither the nomological nor the dependency conception of causation can properly ground mental causation, since they do not deliver the causal efficacy of mentality. Many counterfactualists will dispute this claim and we will elaborate on this shortly. Kim concludes that the relation of causation as dependence, or counterfactual dependence, even if it is a proper and useful causal relation, is not the source of our worries about mental causation. In his view, we need to show that mentality has causal powers to bring about their effects in a continuous process of generation and production.

In his latest writings, he openly sympathizes with Dowe's conserved quantity theory, and seems to suggest that the exclusion argument in fact requires or presupposes this theory, or something very similar (Kim, 2010 p. 236). Quite independently of the difficulties with Dowe's theory in general, which we have discussed in the previous section of this thesis, tying the exclusion argument to this theory of causation is problematic. Dowe himself writes: "…*to suppose that the conserved quantity theory will deal with causation in other branches of science also requires commitment to a fairly thorough going reductionism, since clearly there is nothing in economics or psychology that could pass for a conservation law".* (Dowe, 2008).

Right after saying that, Dowe also reflects on some kind of causal pluralism (such as the view of Nancy Cartwright) as an alternative. Perhaps the conserved quantity theory works at the level of fundamental physics, but a different theory is needed to take care of causation in various higher-level special sciences. In any case, we cannot deny that Dowe's theory directly applies only in the domain of fundamental physics.

This results in a dilemma. We must either presuppose an all-encompassing and strong reductionism, or we must allow the possibility that Dowe's conserved quantity theory simply does not cover mental causation and other higher-level causation (Raatikainen, 2018 p. 40). In the former case, the exclusion argument becomes redundant and the overall argument circular, since the hoped-for conclusion of reductionism is already presupposed (Loewer, 2002). In the latter case, considerations about mental causation and other higher-level causation must be based on a different theory of causation and Kim is wrong in assuming that a theory like Dowe's can and must be applied here.

To conclude, if the exclusion argument actually presupposes a specific theory of causation, i.e. a production view of causation such as Dowe's conserved quantity approach, and that theory in turn presupposes a strong reductive view, it seems pointless and circular to appeal to the exclusion argument in defense of a reductive view. If we assume that the exclusion argument presupposes such a theory of causation, we are in need of another argument, independent from the theory of causation we choose, to support such a strong reductionist view.

The other option would be to develop a separate theory of causation to support mental causation and the higher-level special sciences. But for sure, Dowe's theory applies only at the fundamental physical level and it is not legitimate to take this theory as the general account of

causation including mental causation. The causation-as-production view in the form of Dowe's conserved quantity theory cannot be used to back up the exclusion argument.

Overdetermination & Causation-As-Dependence

Now that we rejected the conception of causation-as-production as suitable for supporting the exclusion argument, let's take a closer look at the counterfactual analysis of the problem.

We have seen as Kim thinks that the counterfactual approach faces serious problems in grounding causation and mental causation in general. We could say that the counterfactual approach to overdetermination face the same kinds of difficulties as the counterfactual approach to causation. The common root of the difficulties is that counterfactual causation is an extrinsic account of causation, while there is a strong intuition that causation is an intrinsic relation. According to the counterfactual approach, whether an event depends on another is affected by extrinsic changes in the circumstances.

The lesson to be learned from the troubles for the counterfactual approach to causation might be that there are in fact two kinds of causation, an intrinsic and an extrinsic view, i.e. production and dependence, rather than that causation is not analyzable in terms of counterfactuals. In a parallel fashion, then, it might be thought that the lesson from the troubles for the counterfactual approach to overdetermination might be that there are in fact two kinds of overdetermination, the production kind and the dependence kind, rather than that overdetermination is not analyzable in terms of counterfactuals (Won, 2014 p. 222). If there are indeed two kinds of overdetermination, here is a plausible response to the overdetermination problem in terms of counterfactuals. It might be highly implausible to hold that every case of mental causation is a case of overdetermination if such overdetermination is thought of as the production kind, i.e. independent overdetermination. A typical example is the firing squad case, where two or more members of a squad simultaneously fire at and 'kill' their targets. The killing of the target is multiply overdetermined, since if all but one have ceased to fire, the target woulg have been still killed. But this is not the kind of overdetermination we have in the mental-physical case. Nonreductive physicalists can hold that mental causation involves multiple dependence causes, and there is nothing objectionable about an effect's being overdetermined by dependence causes. Our intuition against systematic overdetermination stems from considering the situation as one of overdetermination by production causes. "The kind of overdetermination [that nonreductive

physicalism] requires is innocuous when causation is understood in terms of counterfactuals" (Loewer, 2002 p. 661).

The matter though is not that easy. The following problem arises. According to the definition of the dependency conception, in order for something to be a cause of an effect requires that it should make some difference to the effect. On the other hand, in cases of genuine overdetermination the effect should make no difference to the effect. So, it seems that the dependency conception leaves no room for the notion of causal overdetermination. The very notion of overdetermination by dependence causes looks self-contradictory. counterfactualists might adopt this strategy to suggests that the threat of overdetermination does not even arise, if causation is understood in terms of counterfactuals.

Let us agree that the dependency conception of causation cannot give a coherent sense to the idea of overdetermination. If so, however, doesn't that mean that the overdetermination problem presupposes the production conception of causation? Even if overdetermination requires production, it doesn't seem right to think that we can just make the overdetermination problem go away by rejecting the production notion and moving to a dependency conception of cause. Bennett also thinks that it is "wrong to assume that the pure dependence notion alone would dissolve the problem completely" (Bennett, 2008). Bennett writes: "But while I certainly agree that the production view [of causation] is often in the background of discussions of the problem … I do not agree that the problem itself actually requires it. I do not agree that rejecting it makes the issue go away" (Bennett, 2008).

It is indeed possible, that from the perspective of the counterfactual approach (and the dependence notion of causation generally), mental properties seem perfectly suitable for causing physical changes. This observation, though, does not yet make the exclusion problem go away, and it does not really tell us where exactly the exclusion argument goes wrong.

Our worries about overdetermination arise out of very simple considerations. Consider a case in which two events are supposed to be two distinct causes of an effect. Suppose this case does not involve preemption. Suppose further that the two events are not plausibly thought of as jointly causing the effect. Then, some sort of causal redundancy appears to be involved, and the worry about overdetermination arises. It seems that here we are only appealing to a clear and simple

intuition, an intuition involving the distinction between overdetermination and joint causation (Won, 2014 p. 224). The dependence notion of cause does no justice to this distinction, though.

Let's sum up. The overdetermination problem presupposes the intuitive notion of overdetermination. And maybe that notion carries a commitment to a certain conception of causation, i.e. some sort of production conception. It may be said, then, that the overdetermination problem really does depend on the production conception of cause. Does this mean that the problem is resolved when we reject the production conception and move to a dependency conception?

In a sense, yes; if we can really reject the production conception of causation in favor of a dependency conception. There is no problem of overdetermination, if causation is dependence. But there are many challenges for explaining causation in terms of counterfactual dependence, as we have seen. For instance, if a dependency conception can neither make sense of overdetermination nor distinguish it from preemption or joint causation, that is a deficiency of that conception. On the dependency conception, we cannot even sensibly ask whether the effect of a mental cause is overdetermined (Won, 2014 p. 226). It is an open question if by simply appealing to a dependency conception of cause makes our worries about overdetermination go away or rather ignores them. However, the possibility that it might turn out that causation is in fact dependency, and it is rather the very notion of overdetermination that is problematic cannot be yet ruled out.

A possible solution

As we discussed in the previous sections, Kim has argued that counterfactuals depend on laws and regularities, and consequently they share the well-known problems of regularity theories. It is true that regularities may indeed provide us clues for our knowledge of counterfactual conditionals. However, the relation between them is in reality much more complicated than Kim seems to think, since our knowledge of such counterfactual dependencies may be based, for example, on our careful analysis of the underlying mechanism or on active experiments, and so on. Remember from the previous section, that apart from Mackie's initial counterfactual approach, later ones such as Lewis' metaphysics of counterfactuals or Woodward's interventionist counterfactuals, were much more sophisticated and in a great respect they got rid of the regularists' obstacles.

Kim's critique views us as mere passive observers of constant conjunctions, that is, observable regularities. But as researchers, we can also actively examine reality and conduct experiments; by manipulating the world, we can often distinguish misleading regularities that have a preceding common cause, for example, from genuine cases of causation. Assume, let us say, that we have observed a regular connection between certain values of X and Y. According to the regularity theory, this amounts to causation, period. According to the interventionist approach, by contrast, we may begin to examine whether this is, after all, a mere correlation, perhaps due to a common cause, (a certain value of) Z. We can then, for example, keep the value of X fixed and vary the value of Z; if the latter results changes in the value of Y, we can, at least tentatively, conclude that it is Z rather than X that is causally relevant to Y. Consequently, the counterfactual approach, especially the interventionist approach, does not, as Kim thought, collapse into the regularity theory and inherit its problems.

Furthermore, the epistemological question of how we can know whether a particular pair of events is a case of causation and not mere correlation, and the metaphysical question of what it means for a pair of events to count as a case of real causation seem to be two distinct issues. Our knowledge of particular causal relations is empirical and, of course, always revisable by further observations and experiments. Particular causal relations may be even unknowable. But as philosophers, we may nevertheless be able to analyze what is required for c to cause e, in contrast to mere correlation. The latter does not require that we can actually decide in every concrete situation whether it really is a case of genuine causation (Raatikainen, 2018 p. 42).

In sum, Kim's critique of the counterfactual theory focuses only on an intuitive and quite naïve form of the counterfactual approach and it conflates the epistemological with the metaphysical aspect of causation. The difficulties he presents can well be solved by more sophisticated theories, and by the interventionist theory in particular.

Nonreductive physicalists have tried to resolve the exclusion/overdetermination problem through appeal to counterfactuals. Árnadóttir and Crane (Árnadóttir, et al., 2013), for example, refer to Mackie's classic fire example, where a house fire was caused by an electrical shortcircuit, but clearly, the short-circuit is neither necessary nor sufficient for the fire, which could have started for other reasons or not started despite the short-circuit. From the point of view of the causes as difference-makers picture, the short-circuit; the presence of oxygen; the presence of inflammable

material nearby etc. are all relevant difference-making causes of the fire. Moreover, the fire was not overdetermined by them: It is not the case that the fire would still have occurred even if one (or more) of these causes had been absent. Nevertheless, none of these excludes the others as causes; they are all equally causally relevant. The exclusion principle, as it is ordinarily formulated, is thus false: an effect may well have simultaneously several relevant causes without being causally overdetermined.

As to causal overdetermination, the standard way to explain it is in counterfactual terms: the common idea is that an effect E is overdetermined by two causes A and B only if, roughly, E would have occurred even if A had not occurred, and E would have occurred even if B had not occurred (and, finally, E would not have occurred if neither A nor B had occurred). If the interventionist theory is followed, the relevant counterfactual conditionals must be accordingly formulated in terms of it: "*If an intervention had prevented A from occurring, E would still have occurred*". For example, in the case of the firing squad shooting, were the members of the squad fire simultaneously, the victim would have died even if a particular shooter had not fired his gun. Even Kim often explicates overdetermination in terms of counterfactual dependencies. By contrast, it is not too clear how overdetermination should be even defined without any help of counterfactuals.

Let's turn once again to the argument about the two kinds of overdetermination. Funkhouser (Funkhouser, 2002) and Bennett (Bennett, 2008) in particular have developed arguments on this issue and particular on the observation that the relation between the mental and the physical causes of a behavioral effect is much more intimate than in the paradigmatic cases of overdetermination (such as the firing squad example).

Funkhouser (Funkhouser, 2002) distinguishes two different types of overdetermination. Paradigmatic examples of overdetermination, such as the firing squad example, are cases of independent overdetermination, since the overdetermining causes work through separate mechanisms and distinct spatial regions. In cases of the so-called incorporating overdetermination, by contrast, the alleged overdetermining causes share a mechanism or causal pathway and differ only in their properties within this shared mechanism. The case of mental causation with supervenient psychological properties and their subvening physical properties is an example of the latter. They share the same causal pathway. Now a case of independent overdetermination is normally a rare coincidence and it is indeed implausible that such overdetermination would occur systematically.

But, as for incorporating overdetermination, it is necessarily systematic, if non-reductive physicalism is granted.

Bennett (Bennett, 2008), on the other hand, focuses on the standard examples of overdetermination (independent overdetermination), and contends that the holding of the appropriate counterfactuals is at least a necessary condition for causal overdetermination. She adds an important qualification, though. She argues that the relevant counterfactual conditionals must not only be true, but its antecedent must also be metaphysically possible. Standard cases of overdetermination of the firing squad kind satisfy this premise; the occurrence of one overdetermining cause without the others is in all cases metaphysically possible.

But given the supervenience assumption, where P is a physical state, M a mental one and B the behavioural effect, the critical counterfactual

If P had occurred without M, B still would have happened

is true, but the antecedent is not metaphysically possible, since the presence of P entails the presence of M with metaphysical necessity. Consequently, the causal efficacy of mental properties does not in the end lead, in the context of non-reductive physicalism, to overdetermination (in the standard sense) contrary to what the essential step of the exclusion argument claims. Bennett concludes that, assuming non-reductive physicalism and a counterfactual analysis of overdetermination, the exclusion principle is, after all, false (Raatikainen, 2018 p. 36).

Finally, the issue of overdetermination is similarly analysed in the interventionist framework. The interventionist theory, together with the supervenience thesis, make overdetermination problematic (Woodward, 2008). From the interventionist perspective, whether a mental state M and the physical state P realizing it really overdetermine the behavioral effect B requires that we can construct a causal system that includes a variable for both M and P (and their negation). However, this requires that one can vary their values independently of each other (like one could, by a hypothetical intervention, prevent one shooter firing his gun without affecting the others in the paradigmatic firing squad example of overdetermination). Only if an intervention were to change the variable from having M to not having M, while P is kept fixed, could one evaluate the truth or the falsity of the critical counterfactual, and if B would still have occurred (Raatikainen, 2018).

But if the supervenience thesis is true, that is, if the facts of the physical level determine the mental level with metaphysical necessity, constructing the relevant causal system is simply impossible. Consequently, the question of overdetermination does not even make clear sense in this context. According to these philosophers, this gives another reason for doubting the soundness of the whole exclusion argument. Note that the paradigmatic examples of overdetermination (e.g., the firing squad) result in no such problems in the interventionist framework; this suggests again that the inference to the conclusion that M and P do overdetermine the effect B is problematic. All this harmonizes well with the above-mentioned arguments of Funkhouser and Bennett.

Of course, all these arguments concerning overdetermination and the exclusion principle could be disputed further. One could rather interpret the whole thing as showing that M is epiphenomenal. But in any case, it is at least fair to conclude that the exclusion premise is much more problematic and open to controversy than Kim have suggested. At the moment the prospects of the exclusion argument do not look particularly promising. Also, the choice of theory of causation is relevant for the exclusion argument, as can be shown by this discussion.

Concluding Remarks

We have paid close attention to the problem of overdetermination, especially Kim's exclusion argument and we have explored its relation to different views on causation and mental causation. It turns out that the exclusion argument and its premises do not really cohere that well with the counterfactual theory of causation or any causation-as-dependence theory. And if the exclusion problem presupposes a specific theory of causation, such as Dowe's conserved quantity theory or any other mechanistic account with a strong reductive view, then the argument becomes circular and pointless. In that case, the exclusion argument does not really threaten non-reductive views and need a separate argument against them.

One possible explanation for this is that our concept of cause is not univocal, as we have already pointed out several times (Hall, 2004). The dependence notion is guided by the intuition that a cause is something that makes a difference to its effects and is naturally analyzed using counterfactuals. The production notion is guided by the intuition that a cause is something that produces or generates its effects through a chain of events. Such accounts analyze causation as involving the transfer of some physical quantities (Dowe, 2000) (Salmon, 1984). Production and

dependence are typically observed together, but sometimes they are not. Cases of pre-emption, for example, illustrate just how they can come apart, since they exhibit production without dependence.

The overdetermination case also showed us how these two approaches to causation deal with the problem differently. According to the counterfactual account a particular event can depend on many co-occurring events. The motions of one's body, for example, depend counterfactually on one's mental states (what one desires); on one's brain states and on several other states and events. Also, depending on the view of causation we choose there are different kinds of overdetermination. The "overdetermination" involved in mental causation is not like the standard overdetermination case (which takes causation as production). And as we said, if the overdetermination problem rests on the production conception of causation, this makes the whole argument circular.

To conclude, the kind of overdetermination involved in the mental-physical case can be understood as overdetermination by dependence causes, which non-reductivist counterfactualists find unproblematic. A possible picture of mental causation that emerges at this point, given the two distinct notions of causation, is to view mental causation as involving both dependence and production causation, such that the mental cause is a dependence cause of its effect, while the underlying physical cause is a production cause of the effect. The mental and physical causes do not jointly cause the effect in the production version of cause. The exclusion problem is often put this way: what causal work is left for the mental cause to do if the physical cause does all the work? But this question is misleading if we are dealing with different kinds of causation.

Insofar as the overdetermination threat is concerned, this double causation approach is a good enough way of dealing with it. However, this does not mean that this approach works as a universal solution to the mental causation problem. If we accept the double view of causation in respect to mental causation in general, the dilemma between overdetermination and epiphenomenalism still stands, since the mental does not cause the effect in the same way that the physical does. The mental then is still exposed to the charge of epiphenomenalism. In general, once we abandon homogeneity of mental causation and physical causation, we risk the charge that our view leads to epiphenomenalism. One of the original motivations for physicalism is exactly the fact that mental and physical states share the same ontological status and are therefore capable of causing – in the same sense of causing – the same effects.

Reductive Physicalism

With all the difficulties encountered by mental causation and the nature of the relationship between the mental and the physical, as well as the constant threat of epiphenomenalism, one may wonder why we don't just reduce the mental to the physical. If we endorse a reductionist view, then mental causation is just a species of physical causation, and is therefore no more problematic than plain old physical causation; or it is just physical causation described using mental vocabulary.

The theory of causation we would choose in the case of reductionism is the one that best applies to physical causation. This is again a huge debate and part of it we have encountered in our reflection on causal theories in the previous section.

The Causal Status of Special Sciences

Reductive physicalism goes hand in hand with a serious consequence though. Accepting a reductionist view means that the properties expressed by the predicates of a higher-level science, say a psychological theory, are identical to the properties expressed by the predicates of a physical theory. Consequently, there is no autonomy left to the special sciences.

In a reductionist spirit, Oppenheim and Putnam in their "Unity of Science as a Working Hypothesis" (Oppenheim, et al., 1958) divided all of science into levels, starting at the bottom with elementary particles and building up to molecules, cells, individuals, and societies. They argued that the science at each level was reducible to the next lower level, and thus that the laws of micro-physics are the basic laws of all sciences. But some philosophers would insist that this picture is deeply misleading and that higher-level properties enjoy an autonomous causal status.

Fodor, characteristically, reacted against the Unity of Science movement and supported the view that nature has joints at many different levels, so at each level there can be genuine sciences with their own conceptual apparatus, laws and explanations, emphasizing the autonomy and irreducibility of the special sciences. He has argued that the multiple realizability argument "refutes psychophysical reductionism once and for all" (Fodor, 1974). Fodor argues that the dream of a single all-encompassing unified science that could be used to explain everything could never exist.

Certain sciences like psychology, he argues, are "autonomous" in that they investigate realms governed by unique special laws. Psychological laws, he argues, could be realized radically differently in numerous physical systems, in ways that make it impossible to derive psychological generalization from the general laws of physics. This is true, he argues, even if minds are ultimately made of merely physical substances. Despite his acceptance of physicalism, he writes: "… I am strongly inclined to think that psychological states are multiply realized and this fact refutes psychophysical reductionism once and for all" (Fodor, 1998 p. 9).

According to the supporters of the Autonomy Thesis (the thesis that argues for the explanatory and causal autonomy of the Special Sciences), psychological explanations typically abstract away from details of lower-level implementation, appealing instead to their own distinctive kinds and laws. If the structure of the causal order reflects these explanatory practices, mental properties need not be threatened by exclusion by the physical ones. Mental and physical causes can peacefully coexist (Woodward, 2008).

The Thesis supporting the Autonomy of the Special Sciences,, in the sense that higher-level properties have causal powers independent of those of their more basic physical properties, was initially supported by the emergentists. Emergentists claimed that systems with a high level of organizational complexity have causal powers that emerge from their constituent elements and are distinct from them.

Here, we will discuss the causal autonomy of the special sciences in the context of nonreductive physicalism rather than emergentism, since non-reductive physicalism is a more familiar framework for our purposes. This framework has been nevertheless criticized by several philosophers, who have argued that its commitment to the causal efficacy of higher-level properties makes it inherently unstable. In this final section, we will try to offer a defence and an opposition of this thesis of the causal autonomy of the special sciences in terms of different accounts of causation.

The Causal Status of Special Sciences and Non-reductive Physicalism

What does it mean to say that special-science properties, in particular mental properties, have causal powers that are independent of those of their physical realizers[1]? What is involved in

1

affirming or denying the claim that certain special-science properties, like mental properties, are autonomous and independent of those of their physical realizers?

The following thesis captures the view of many philosophers who deny the causal autonomy of mental properties and endorse reductionism (Menzies, et al., 2010).

For all mental properties M and physical properties P, if an instance of property M is realized by an instance of property P, then the causal powers of the M-instance are a subset of the causal powers of the P-instance.

The formulation of this thesis in terms of subsets allows also for the special case in which the causal powers of the mental state are identical to those of its realizing physical state. So, for example, the instances of M and P may have the exact same causal powers to produce the behavioural effects B1 and B2. But equally, the causal powers of the instance of M may be a proper subset of those of the corresponding instance of P: perhaps the instance of M has the power to produce B1 and B2 under certain conditions, while the instance of P has the power to produce B1, B2 and B3 under the same conditions (Menzies, et al., 2010 p. 110).

Now the denial of the above thesis means the assertion of the causal autonomy of mental properties. This means that according to the Causal Autonomy Thesis, some mental states have causal powers that are not causal powers of their realizing physical states. This claim is controversial and denied by many philosophers, such as (Kim, 1998) and Sydney Shoemaker (Shoemaker, 2001). They state that the causal powers of a mental state must be identical with those of the physical state that realizes it. There is actually a broad acceptance of the reductionist thesis in philosophy of mind, which, possibly, stems from a more general preference for lower-level, physical causal variables over higher-level, special-science variables. It is the physical properties and states that do all the causal work, it is assumed, and properties and states supervening on them derive whatever causal efficacy they have from the underlying physical ones.

Special Sciences & Counterfactual Causation

In this section we will explore if the difference-making account of causation, i.e. counterfactual causation, can determine whether the Reductionist Thesis described above is true or

not. Woodward's interventionist theory is a paradigmatic difference-making account of causation. On this theory, the causal relata are variables, and causation relates changes in one variable to those in another. A variable X is said to cause variable Y iff after an intervention on X the value of Y changes too. In the case of mental causation the theory uses variables whose values represent the occurrence or non-occurrence of an event, or the instantiation or non-instantiation of a property. The central difference between correlations and causal relations according to the interventionist account is that a genuine causal relation is stable under interventions that change the values of the cause variable.

Recall that the reductionist thesis states that if an instance of a mental property M is realized by an instance of a physical property P, then the causal powers of the M-instance are either identical to those of the P-instance or a subset of it. If this thesis is true, then its negation, the thesis of the Causal Autonomy of Mental States, must be false and vice versa.

It is useful to evaluate the thesis by considering a schematic example. Suppose you have an intention to open the window (an instance of mental property M) and you raise your arm (an instance of behavioural property B); and suppose that this intention (the M-instance) is realized by some neural state (an instance of neural property N1), but it could also have been realized by other neural states (say, instances of the neural properties N2, …, Nn).

What is the cause of you raising your arm – the mental state M or the neural one N1? Following the counterfactual theory of causation, it can be the case that the mental state M, and not the neural state N1, is the cause of the behavioral outcome B. Consider the following counterfactual conditionals (Menzies, et al., 2010). The pair (1a) and (2a) are counterfactual statements of assertion, while (1b) and (2b) are stating the absence of the mental or the neural state.

(1a) M □→ B

(1b) ~M □→ ~B

(2a) N1 □→ B

(2b) ~N1 □→ ~B

Thus it may be the case that both counterfactuals (1a) and (1b) are true, but not both counterfactuals (2a) and (2b) are true, since you raising your arm rather than not raising it was caused by you having the mental property M rather than not having it, and not by you having the neural property N1 rather than not having it. If we accept multiple realization, there are many common situations in which a supervenient state has causal powers not possessed by the subvenient state that realizes it.

The two pairs of counterfactuals (a & b-counterfactuals) express the contrastive character of causation. It is increasingly common to think that causal judgments in fact involve (even if often implicitly) a class of contrasts for both cause and effect—that they contrast the putative cause and effect with some alternatives. Different choices of contrasts for the same cause and effect result in different causal judgments, some of which may be true, some false.

In our case, the function of these counterfactuals is to ensure that the potential causes are of the right degree of specificity; that they are neither too specific nor too broad to account for the difference made to the effect. A-counterfactuals rule out causes that are not specific enough, while b-counterfactuals rule out candidate causes that are too specific.

Returning to the example about you raising your arm with the intention of opening a window, we can imagine conducting experiments the results of which confirm (1a) and (1b) and disconfirm (2b). Such experimental evidence would establish the falsity of the Reductionism Thesis, since it would establish that a mental state has a causal power not enjoyed by the physical state that realizes it. Thereby the truth of the Causal Autonomy thesis can be demonstrated.

Suppose, however, that we want to study and explain changes of your fine-grained motor control rather than your overall behavioral output. This change might be explicable only in terms of a variation in your neural states and not in terms of a variation in your mental states. In this case a neural state would be the difference-making cause of the effect. Generally, the right level of causation is determined by the contrast to be explained and by the empirical facts about which variables can be varied in such a way as to account for the given contrast.

In the previous example with the window opening, the higher-level variable is the source of causal influence. But in other circumstances lower-level variables can constitute the right level of causation. Facts about the world determine, which is the right level of causation. Empirical facts

determine which type of variable, a higher- or a lower-level one, is such that variation in its value can bring about a variation in an effect variable.

Moreover, we have to admit that in order for the higher-level and not the lower-level variable to be the source of the causal influence, the case in hand has to be a multiple-realizability case, where the causal relation between the mental state and behavior holds, even if the physical realizer changes. In realization-sensitive cases (Menzies, et al., 2010), though, the Reductionism Thesis is true, since the behavior is caused by the mental state, which is identical to the physical state. If the realizer of the mental state changes even slightly, the behavioral effect also changes and the relation between the mental state and the behavior does not stand. In other words, if M were realized by any neural state other than N1 (such as N2, N3, and so on), then B would cease to hold.

One benefit of formulating the difference-making conception of causation in terms of counterfactuals is that it makes this question logically tractable. Returning to the case of the mental state M, the neural state N1 and the behavior B, one can prove that the causal powers of M are a subset of those of N1 only under very special conditions; where multiple realizability is not the case (in realizability-sensitive cases).

When might we expect these conditions to obtain? If the mental property M were identical to the neural property N1, then we would certainly expect instances of M to stand in realization-sensitive causal relations with respect to instances of N1. The fact that M-instances had certain effects when and only when N1-instances are present would simply reflect the identity of the properties. It is hard to think of any explanation other than the identity of the properties. But this explanation will not be available if we assume that the higher-level properties are multiply realized by physical properties, and so no identical with them.

This result is of great significance with respect to the question of whether the Reductionism Thesis or the Causal Autonomy thesis is true. If some mental state M is multiply realizable, then this mental state has a causal power to bring about a certain effect that does not belong to the set of causal powers of its physical realizing state. In these cases, the Reductionism Thesis is false and the Causal Autonomy Thesis true.

What is the upshot of this discussion? If we have reason to believe that a mental state is multiply realizable and so stands in a realization-insensitive causal relation to some other state, then

we are entitled to think that this higher-level causal relation is independent of any lower-level causal relation enjoyed by the physical/neural realizer of the mental state. We have plenty of reason to believe that mental states do indeed stand in realization-insensitive causal relations to other states. Given that a mental state is typically realized in many different ways, we can expect that whatever causal powers it has, it has then independently of the particular way it is realized. In other words, we can expect that a mental state's causal powers do not depend on which of its possible realizers happens to be the actual one. Generally, there is reason to think that most higher-level causal relations are multiply realizable in ways that ensure their autonomy. In the special sciences causal relations are typically required to be invariant under changes to the physical realization of the higher-level properties involved. What is implied by our discussion so far is that the macro-explanation (higher-level) conveys essential contrastive, difference-making information that is not conveyed by the microstate (lower-level) explanation.

To conclude, given the contrastive character of causation, we can establish the right level of causation and justify - in multiple realizability cases – the irreducibility of higher-level causes to physical causes.

Special Sciences & Mechanistic Approach

Another possible defense of the thesis of the causal autonomy of the special sciences can be found in the mechanistic account of Stuart Glennan, which we also encountered in the previous section on causal theories. Salmon's and Dowe's accounts aim at a mechanistic theory of physical causation, but their accounts of conserved quantity exchange is too narrow to describe cases of causation among higher-level entities. We would surely miss something if we tried to explain a complex behavior, in terms of the exchange of conserved quantities.

Glennan's account, on the other hand, is broader than this of Salmon and Dowe and explains mechanisms at levels higher than fundamental physics. He argues for the metaphysical supervenience of the interactions involved in higher-level mechanisms on fundamental physical interactions, but he does not think it is helpful to describe these higher-level mechanisms in terms of their subvenient parts. Higher-level interactions form higher-level kinds.

In the same way, Glennan also argues for a distinction between fundamental laws of physics and mechanically explicable laws, which are laws "explained by the behavior of some mechanism" (Glennan, 1996 p. 62). Special sciences are characterized by mechanically explicable laws and not by the fundamental laws of physics (Glennan, 1996 p. 50). Non-fundamental mechanically explicable laws are realized by fundamental laws, but are not reduced to them. The fundamental laws govern the interactions of the parts of the mechanism, which is, however, more than the sum of its parts. The mechanism has explanatory and epistemic autonomy.

In conclusion, both from a counterfactual and a mechanistic perspective, is the Autonomy of the Special Sciences to be defended. However, there is a huge difference between the two approaches in terms of their epistemological strength. According to the mechanistic account of causation, we need to understand and explain the underlying mechanism in order to understand higher-level causal relations, since the later are grounded in the former. But identifying the mechanism is not always a possible task and causal inference depends on counterfactual reasoning and interventions, as causal model theories show.

In reality, science functions by combining both counterfactual reasoning and careful explorations of the underlying mechanisms. A well-defined hypothesis about an underlying mechanism is informed by counterfactual reasoning and inference, but the knowledge of a mechanism can in turn inform the methodology of causal inference (the choice of the variables to be controlled or to be intervened upon). A full understanding of a causal relation requires both counterfactuals and mechanisms, both dependence relations and relations of generation and production.

Conclusion

The present thesis was an attempt to combine two major branches of philosophy; philosophy of mind and philosophy of science. Minds are part of this world and simultaneously the means to grasp, perceive and feel the world. This double role of the mind, as being-of-the-world and being-into-the-world makes it an extremely complicated object of research. The initial thought which inspired this thesis though, is to explore how philosophy of science, which is mainly concerned with the assumptions, foundations, methods and implications of science could serve as a

foundation of the understanding and explanation of the mind and the mental in general. We have focused on a particular topic of philosophy of science, the study of causation, and our intention was to use different approaches to causation and investigate how they inform the debate on mental causation. What does it mean for two events to stand in a causal relation and how do different concepts of causation result to different readings of mental causation issues?

Regarding the different accounts and concepts of causation, I have focused on the broad distinction between causation-as-dependence and causation-as-production. Whether there are indeed two distinct, coexisting concepts of causation or there has to be a univocal concept, is a question that goes beyond the scope of this discussion. In order to narrow the issues of mental causation, on the other hand, I have selectively focused on the debate between reductionism and non-reductive accounts of the mental, since I strongly believe that the side one chooses on this debate defines the features of the causal theory that best defends it. A different account of causation is needed, if we are to defend a radical reductionist thesis than the one appropriate to ground the supervenience thesis.

Following these general considerations, this thesis has shown that the mechanistic account of Salmon and Dowe is too narrow to explain and describe complex phenomena such as mental events, unless we accept radical top-down reductionism. This theory can only be applicable at the most basic physical level and as it was shown in the section *The Problem of Overdetermination & Causal Theories* it actually presupposes reductionism.

Higher-level physical phenomena, such as mental events, require for their explanation a causal theory, which offers an account of the relation between different levels, from the fundamental physical level to higher-order psycho-physical phenomena. The overall conclusion is that both sophisticated counterfactual accounts of causation, such as the interventionist thesis, as well as mechanistic accounts, which explore causal mechanisms at higher levels of physicality than that of the fundamental micro-physics, such as Glennan's account, are suitable for explaining mental causation.

I consider it as a prerequisite for the defending of any non-reductive thesis to also hold an account of causation, which attempts to bridge the different levels of physical reality. If this is a causation-as-dependence theory or a causation-as-production one, is not at this point relevant.

Counterfactual accounts of causation are mostly epistemologically-driven. They take as their point of departure the inferential possibilities we have and define causation according to them; whereas mechanistic accounts are driven by the assumption that causation lies in the intrinsic nature of the relata; they are therefore metaphysically-driven. These different points of departure create tensions between the different causal accounts, as it is made clear by our discussion in the Section *Causal Theories*.

As we said, both families of causation (dependence and production) can give some explanation of mental causation. The final debate, which stems from this above-mentioned tension between the accounts, is whether causation understood as counterfactual dependency or as invariance under interventions is more fundamental than causation as a mechanism of production and vice versa. Counterfactual reasoning and interventions are necessary as a causal inference tool. Whether there is a hidden or a detectable mechanism at work, which actually offers these dependence relations its truth-conditions, is a one-million-dollar question. The only way to progress is to keep both of these concepts in parallel and make them work together. Start from a mechanistic hypothesis, infer through hypothetical or actual counterfactuals a relation of dependence and then according to the findings, evaluate you hypothesis and search further for a possible mechanism.

The final word of this thesis concerns some future perspectives on these issues. It would be a very inspiring topic for further analysis and discussion to link our contemporary methodology in the special sciences with the different theories of causation and try to detect in each case which causal theory is presupposed or necessary. In this way, a broader perspective could be gained, as of how we can understand causation in the special sciences and if this understanding is helpful or not.

Since my scientific background comes from cognitive science, a careful analysis of which concept of causation is presupposed when neuroscientists link a behavior (or a mental state) with a neural state, could shed further light to this discussion and create further links between philosophy and neuroscience. There are several ways to link neural activations to behavior; for example through distributed activations and results from fMRI or EEG, or through lesions in specific brain areas or even through the investigation of neuron-level mechanisms, such as protein activation, neurotransmitter quantity etc. Each one of these methods indicates to different concepts of what this link between the behaviour (outcome) and the neural state (the base) is. I strongly believe that

by putting both of our causal concepts to work in parallel, science and philosophy can inform each other and promote greater scientific and epistemological progress.

Bibliography

Árnadóttir Steinvör Thöll and Crane Tim There is No Exclusion Problem [Book Section] // Mental Causation and Ontology. - Oxford : Oxford University Press, 2013.

Bennett Jonathan Correspondence between Descartes and Princess Elisabeth [Book]. - 2017.

Bennett Karren Exclusion Again [Book Section] // In Being Reduced: New Essays on Reduction, Explanation, and Causation. - [s.l.] : Oxford University Press, 2008.

Block Ned Anti-Reductionism Slaps Back [Journal] // Mind, Causation, World, Philosophical Perspectives. - 1997. - Vol. 11. - pp. 107-133.

Carnap Rudolf An Introduction to the Philosophy of Science [Book]. - New York : Basic Books, 1974.

Cartwright Nancy Measuring Causes: Invariance, Modularity and the Causal Markov Condition [Journal]. - 2000.

Cheng C-K. The Optimal Partitioning of Networks [Journal]. - [s.l.] : Networks, 1992. - 3 : Vol. 22. - pp. 297-315.

Davidson Donald Essays on Actions and Events [Book]. - Oxford : Oxford University Press, 2001.

Davidson Donald Mental Events [Book Section] // Essays on Actions and Events. - 1970.

Davidson Donald Problems in the Explanation of Action [Book Section] // Problems of Rationality. - Oxford : Clarendon Press, 2004.

Davidson Donald Thinking Causes [Book Section] // Mental Causation. - Oxford : Clarendon Press, 1995.

Davidson Donald Truth, Language and History [Book]. - Oxford : Clarendon Press, 2005.

Descartes Rene Meditations on First Philosophy, The Philosophical Writings of Descartes [Book]. - Cambridge : Cambridge University Press, 1985, 1641. - Vol. II.

Dowe Phil [Online] // Stanford Encyclopedia of Philosophy. - 2008. - June 05, 2020. - https://stanford.library.sydney.edu.au/archives/sum2010/entries/causation-process/.

Dowe Phil Physical Causation [Journal]. - Cambridge : Cambridge University Press, 2000.

Fodor J.A. Special Sciences (Or: The Disunity of Science as a Working Hypothesis) [Journal]. - 1974. - 2 : Vol. 28. - pp. 97-115.

Fodor J.A. Special Sciences: Still autonomous after all these years [Journal]. - [s.l.] : Critical condition, MIT Press, 1998. - pp. 9-24.

Fodor Jerry Explanations in Psychology [Journal] // Philosophy in America. - 1965. - pp. 161-179.

Fodor Jerry You Can Fool Some of the People All of the Time, Everything Else Being Equal: Hedged Laws and Psychological Explanations [Book Section] // Mind. - [s.l.] : New Series 100, 1991.

Funkhouser Eric Three Varieties of Causal Overdetermination [Journal]. - [s.l.] : Pacific Philosophical Quarterly, 2002. - Vol. 83. - pp. 335–351.

Glennan Stuart Mechanisms and the Nature of Causation [Journal]. - [s.l.] : Erkenntnis, 1996. - Vol. 44. - pp. 49-71.

Glennan Stuart Rethinking Mechanistic Explanation [Journal]. - [s.l.] : Philosophy of Science, 2002. - 3 : Vol. 69. - pp. 342-353.

Hall Ned Two Concepts of Causation [Book Section] // Causation and Counterfactuals. - [s.l.] : MIT Press, 2004.

Hausman D.M. and Woodward J Independence, Invariance and the Causal Markov Condition [Journal]. - [s.l.] : The British Journal for the Philosophy of Science, 1999. - pp. 521-583.

Hitchcock Christopher Discussion: Salmon on Explanatory Relevance [Journal]. - [s.l.] : Philosophy of Science, 1995. - Vol. 62. - pp. 304-320.

Horwich Paul Asymmetries in Time [Book]. - Cambridge : MIT Press, 1987.

Hume David A Treatise of Human Nature [Book]. - Oxford : Clarendon Press, 1739.

Hume David An Enquiry Concerning Human Understanding [Book Section] // Enquiries Concerning Human Understanding and Concerning the Principles of Morals. - Oxford : Clarendon Press, 1748.

Jaegwon Kim Mental Causation [Book Section] // The Oxford Handbook of Philosophy of Mind / book auth. Beckermann Ansgar, McLaughlin Brian P. and Walter Sven. - [s.l.] : Oxford Hanbooks Online, 2009.

Kim Jaegwon Can Superveniece and 'Non-Strict Laws' Save Anomalous Monism? [Book Section] // Mental Causation. - Oxford : Clarendon Press, 1995.

Kim Jaegwon Causes and Counterfactuals [Journal]. - [s.l.] : Journal of Philosophy, 1973. - 17 : Vol. 70. - pp. 570-572.

Kim Jaegwon Essays in the Metaphysics of Mind [Book]. - Oxford : Oxford University Press, 2010.

Kim Jaegwon Mental Causation [Book Section] // The Oxford Handbook of Philosophy of Mind. - [s.l.] : Oxford University Press, 2009.

Kim Jaegwon Mind in a Physical World [Book]. - [s.l.] : The MIT Press, 1998.

Kim Jaegwon Physicalism, or Something Near Enough [Book]. - [s.l.] : Princeton University Press, 2005.

Kim Jaegwon Supervenience and Causal Closure: Two Exclusion Arguments and How in the End They Reduce to One [Journal]. - [s.l.] : Korean Journal of Philosophy, 2014.

Kim Jaegwon The Layered Model: Metaphysical Considerations [Journal]. - [s.l.] : Philosophical Explorations, 2002. - 1 : Vol. 5.

Lewis David Counterfactuals [Book]. - Cambridge : Harvard University Press, 1973.

Lewis David Philosophical Papers [Book]. - Oxford : Oxford University Press, 1986.

Loewer Barry Comments on Jaegwon Kim’s Mind in a Physical World [Journal]. - [s.l.] : Philosophy and Phenomenological Research, 2002. - Vol. 65. - pp. 655-662.

Mackie John L. The Cement of the Universe: A Study of Causation [Book]. - Oxford : Clarendon Press, 1974.

Menzies P. and Price H. Causation as a Secondary Quality [Journal]. - [s.l.] : British Journal For the Philosophy of Science, 1993. - pp. 187-203.

Menzies Peter and List Christian The causal autonomy of the special sciences [Book Section] // Emergence in Mind. - Oxford : Oxford University Press, 2010.

Oppenheim Paul and Putnam Hilary Unity of Science as a Working Hypothesis [Journal]. - [s.l.] : University of Minnesota Press, Minneapolis, 1958. - Vol. 2.

Papineau David Thinking about Consciousness [Book]. - Oxford : Oxford University Press, 2002.

Psillos Stathis A Glimpse of the Secret Connexion: Harmonizing Mechanisms with Counterfactuals [Journal]. - [s.l.] : Perspectives on Science, 2004. - 3 : Vol. 12.

Psillos Stathis Causation & Explanation [Book]. - Stocksfield : Acumen Publishing, 2002.

Putnam Hillary The Nature of Mental States [Journal] // Art, Mind, and Religion. - [s.l.] : Pittsburgh University Press, 1965. - pp. 1-223.

Raatikainen Panu Kim on Causation and Mental Causation [Journal]. - [s.l.] : Electronic Journal for Philosophy, 2018. - 2 : Vol. 25. - pp. 22-47.

Reichenbach Hans The Direction of Time [Book]. - [s.l.] : Dover Publications, 1956.

Salmon David Causality and Explanation [Book]. - Oxford : Oxford University Press, 1997.

Salmon David Scientific Explanation and the Causal Structure of the World [Book]. - Princeton : Princeton University Press, 1984.

Salmon David The Causal Structure of the World [Journal]. - [s.l.] : Metatheoria, 2010. - 1 : Vol. 1.

Shoemaker Sydney Realization and Mental Causation [Book Section] // Physicalism and Its Discontents. - [s.l.] : Cambridge University Press, 2001.

Sperry R.W. A Modified Concept of Consciousness. [Book Section] // Psychological Review. - [s.l.] : American Psychological Association, 1969.

Strawson P.F. Causation and Explanation [Book Section] // Essays on Davidson: Actions and Events. - Oxford : Clarendon Press, 1985.

Sur Mriganka C-fos Expression and Accelerated Visual Cued Fear Conditioning in Mice with Visual Input Directed to the Auditory Thalamus. [Journal]. - [s.l.] : Journal of Vision, 2003. - 9 : Vol. 3.

Won Chiwook Overdetermination, Counterfactuals, and Mental Causation [Journal]. - [s.l.] : Philosophical Review, 2014. - 2 : Vol. 123. - pp. 205-229.

Woodward James Explanation and Invariance in the Special Sciences [Journal]. - [s.l.] : The British Journal for the Philosophy of Science, 2000. - pp. 197-254.

Woodward James Making Things Happen: A Theory of Causal Explanation [Book]. - Oxford : Oxford University Press, 2003.

Woodward James Mental Causation and Neural Mechanisms [Book Section] // Being Reduced—New Essays on Reduction, Explanation and Causation. - [s.l.] : Oxford University Press, 2008.