Abstracts Causality in the Sciences of the Mind and Brain, Aarhus University June 27-29th 2016

Bonicalzi, Sofia: What counts as mental causation? Between philosophy of mind and neuroscience

The problem of mental causation lies at the core of the interaction between neuroscience and philosophy of mind. However, one might doubt whether the two disciplines are dealing with the same conceptual issues. In philosophy of mind, the interest for mental causation often grows from the dispute between compatibilist and incompatibilist perspectives about metaphysical determinism and free will. In neuroscience, mental causation remains a central issue, but the focus is shifted from metaphysical determinism to the neurobiological mechanisms regulating the structure of voluntary actions. Here, the problem is not if the acts of the will are or are not deterministically caused by some previous event, but if the will can cause anything. Indeed, modern neuroscience is committed to reject the picture of the will as a force generating a physical movement, which would be at odd with a materialistic conception of the mind. Claiming that consciousness of a decision, where reduced to a biochemical afterthought, does not play a causal role in the production of the subsequent action might signify that conscious will and conscious mental states are just bypassed by unconscious stimuli. Following this line, Wegner has questioned if intentions might be located among the causes of the correspondent actions (Wegner 2004). Within this framework, our sense of causing and controlling events in the external world (sense of agency) might be deceptive. In this paper, I explore whether different explanations of human sense of agency and of the effectiveness of mental causation could be viable. As an example, Haggard et al. introduced the intentional binding effect as a marker of the sense of agency, reflecting a temporal association between intentional actions and correspondent outcomes (temporal attraction of the action towards the effect and of the effect towards the action) (Haggard et al. 2002). Interestingly, intentional binding might be produced by the same circuits, in the frontal lobe, that cause intentional action. An important consequence would be that the same circuits causing the preparation and the initiation of voluntary actions also produce a prediction of the expected outcomes: our sense of agency might be not just a retrospective confabulation, but a "measurable signal within the motor system" (Haggard 2014, p. 883). Within this perspective, the crucial point becomes how to understand what kind of neural mechanisms enable our consciously willed actions, giving individuals the sufficient degree of control that is necessary for causally interacting with the external world. I will argue that, moving beyond the classic compatibilist-incompatibilist divide, also philosophical discussion could highly benefitted from such move.

Colombo, Matteo and Weinberger, Naftali: Modes of connectivity and causality in the brain

Connections between brain components can be studied at different spatial and temporal scales, with different instruments, and on the basis of different data. An important distinction in the study of brain connectivity is that between structural, functional, and effective connectivity. Structural connectivity refers to physical connections linking brain components. Functional connectivity refers to statistical relationships between activities of different brain components. Effective connectivity refers to causal relations between brain components, and it "may be viewed as the union of structural and functional connectivity" (Sporns 2007).

The relationships between structural, functional, and effective connectivity represents a significant challenge to understanding causal interactions in the brain. Each type of connectivity should ideally provide information relevant to determining causal relations between different brain components, at different spatial and temporal scales. In practice, it is far from obvious that the multiple ways by which structural, functional, and effective connectivity are measured will lead to coherent conclusions about causality. In particular, it is unclear how information about structural and functional connectivity should be relevant to determining effective connectivity.

Our paper aims to clarify the various forms of brain connectivity, and their relationships to causal inference. In particular, we ask the question: How does information about structure contribute to and constrain claims about effective connectivity? We argue that present discussions have underplayed the role of structure in one respect, while overemphasizing it in another. Too little attention has been paid to developing an account of how structural knowledge constrains causal claims. At the same time, the common misconception that without such an account we can never establish causal relationships leads to the neglect of existing methods for measuring functional connectivity, which may themselves aid in establishing causal relations.

Although it has been suggested that structural connectivity places "constraints on prior beliefs about effective connectivity" (Friston 2011: 16), information regarding the physical mapping of the brain has played a minimal role in present causal modeling methods for the brain. Certain methods relying on time-series data, such as Granger causality, transfer entropy, and algorithms for causal search, omit structural information by design. In contrast, dynamic causal models provide an explicit link between the functional relationships among neural regions and the physical processes underlying these relationships. Nevertheless, we raise doubts about the degree to which present applications of such models are informed by knowledge of structural connections.

Many discussions of connectivity presuppose that it is impossible to establish the presence of causal relations in the absence of structural information. Yet there are several principles for inferring causality from probabilities without explicit appeal

to such information (e.g. Causal Markov Condition, faithfulness, minimality, etc). There has been little work on the applicability of these principles in the domain of neuroscience. Even if these principles accurately describe the relationship between causal and functional dependence, successfully using them requires an adequate measure of functional dependence. Even so, the existence of such principles undermines the common position that statistical knowledge may only be transformed into causal knowledge by adding information about structural connections.

De Vreese, Leen: Choosing for a descriptive or a theoretical approach to psychiatric classification in the DSM: it is all about the causes.

Starting with the third edition of the DSM in 1980 (The diagnostic and statistical manual of mental disorders III), a descriptive approach to mental disorders was taken as a starting point. The editors were following the point of view much earlier defended by the well-known German psychiatrist Emil Kraepelin (1856-1926). He argued that, unfortunately, a precise description of psychopathology was the best we could have in psychiatry, since too little was known about the workings of the brain. Things did not change too much over the course of several new editions of the DSM, till the members of the DSM-V taskforce promised to move away from the purely descriptive approach towards a theoretical approach which was informed by basic sciences such as genetics and neuroscience. This would also help to minimalize the perverse effects of the descriptive approach, such as the increasing trend of medicalization. However, the result was disappointing for those who had hoped for a big revolution and DSM V is still to a large extent building further on the descriptive approach of its predecessors. The direction in which the DSM should evolve in the future is now in the middle of the philosophical (and broader) debate about the DSM. Some argue that, although (neuro)sciences evolved a lot in the last couple of decades, the brain is so complex that we are still far away from reaching the area in which we can replace the descriptive approach by a theoretical one. Others are much more optimistic and argue that it is already possible for the DSM to start taking the route towards a fully theoretical approach. These convictions are related to convictions about our knowledge of the causes of mental disorders. Some argue that for several mental disorders, we have now gathered enough scientific knowledge to support the idea that they have stable underlying biological mechanisms, others argue that our knowledge about biological causes is far too limited, and/or that mental disorders cannot be scientifically understood from a reductionist point of view, but will always need a biopsychosocial approach. Interestingly, since the "causal knowledge" we have is largely based on psychopharmacological findings, this is also related to the question of what kind of causal evidence we get from psychopharmacology and whether this is a reliable source for knowledge on the causes of mental disorders. In my talk, I want to analyze the current philosophical debate and sort out the relations between convictions about the approach that the DSM should take in the (near) future (descriptive or theoretical) and convictions about the role of biological

causes in mental disorders, about the extent to which knowledge about biological causes have or have not been found, and about what kind of causal evidence we need for this. This way, I will explore whether or not this battle between descriptivists and theoreticists cán be fought out, and whether we really have a choice to make.

Drozdzewska, Anna What do interventions tell us about causation? The dialogue between philosophy and neuroscience.

In this presentation we will focus our attention on interventionism, an approach to causation that grasps the core of experimentation, that of manipulation. The approach was famously developed by, among others, Menzies and Price (1983) and Woodward (2003). Although diverse formulations of this argument differ in specifics, the main assumption across all of them is that causes can be seen as tools through which we manipulate effects. In other words, if an event X is a cause of an event Y, the manipulation of X will change Y. One of the initial goals of the approach was to provide a solution to the Causal Exclusion Argument, made famous by Jaegwon Kim (1998, 2005).

We will discuss interventionism in the light of the ongoing discussions on free will, encompassing both philosophy as well as neuroscience. The ideas presented here were initially inspired by the realization that interventionism, although applied in both domains in a very similar manner, leads to contradicting results. We argue that, for free will to be possible, mental events have to be causally efficacious, which makes the question on the possibility of free will's existence partially dependent on the solution of the Causal Exclusion Argument. The core idea of interventionism follows a line of thinking very broadly applied in scientific experiments, where we manipulate different variables and observe how they influences the effect. Partially, this is also true about the experiments on free will, starting from the classic Libet's experiment (1983, 1985) through the modernized variations conducted by Haggard (e.g. 2001) and Soon (2008). Although neither intentions nor brain processes are manipulated in those experiments, the underlying assumption is that if we know what causes our actions, brain processes or conscious intentions, we might be able to manipulate it. Other experiments try directly to manipulate various parts of the brain, resulting in for instance, movements not initiated by an individual. In neuroscience the interventionist approach leads to conclusion that free will does not exist. In this presentation we will discuss some of the issues with this approach, directly related to the idea of manipulation.

On the other hand in philosophy, interventionism is often used to defend mental causation, and through it, the possibility of free will (List and Menzies, forthcoming). Through the principle of multiple realization connected with the idea of manipulation of the cause, the authors conclude that, roughly, since intentions could have supervened on a different physical event, and if we would have manipulated the intention, the action would not have been executed. It

follows that the cause which, when manipulated, changes the action is the intention and not its (multiply realizable) physical basis. In this presentation we will point out that because of the principle of supervenience, interventionism, when applied to mental causation, proves the inefficacy of the mental rather than the opposite, in line with neuroscientific conclusions. We will conclude by showing that the idea of manipulable causes only works when applied to two events not related through supervenience.

Gebharter, Alexander: Causal exclusion and causal Bayes nets

Causal exclusion arguments, most famously advanced by Kim (2005), can be used as arguments for epiphenomenalism or as arguments against non-reductive physicalism. Epiphenomenalism is the view that mental events are caused by physical events, but have no effects upon any physical events. Non-reductive physicalism, on the other hand, basically consists of three assumptions: Mental properties supervene on physical properties, mental properties cannot be reduced to physical properties, and mental properties are causally efficacious.

In a nutshell, causal exclusion arguments assume non-reductive physicalism and conclude from several premises that mental properties supervening on physical properties cannot cause physical or other mental properties. The notion of causation used in these arguments is, however, typically somewhat vague and not specified in detail. In this talk, I model two variants of the causal exclusion argument by means of causal Bayes nets (CBNs). The hope is that such a modeling allows to draw new and empirically informed consequences about the validity of causal exclusion arguments, which would be directly relevant for the sciences of the mind and brain.

The talk is structured as follows: In the first part I briefly introduce two variants of the causal exclusion argument. In the second part I reconstruct these two variants within the theory of CBNs and evaluate their validity. This requires an answer to the question of how supervenience relationships should be represented in CBNs and a test for evaluating whether the instantiation of a property X at least sometimes contributes something to the occurrence of another property Y. I will argue that supervenience relationships can be treated similar to a CBN's causal arrows. I will then suggest a method for testing a property's causal efficacy on another property. This method is based on one of the theory of CBN's core axioms, viz. the causal minimality condition (cf. Spirtes et al., 2000, p. 31). I conclude by demonstrating that mental properties supervening on physical properties cannot be causally effective if two assumption typically made when doing causal modeling are met, viz., the causal Markov condition (cf. Spirtes et al., 2000, p. 29) and the causal minimality condition. In the third part of the talk I investigate the consequences of these findings for the interventionist debate on the causal exclusion argument (cf. Shapiro & Sober, 2007; Woodward, 2008; Baumgartner, 2010).

Gervais, Raoul: Causality and Evidence in Psychology
When it comes to causal explanation, two approaches dominate psychology: the cognitive approach, which seeks to explain behavior in as effects of mental constructs (cf. Lamberts & Goldstone 2004 for reviews), and the functional approach, which seeks to explain behavior as effects of elements in the environment (Chiesa 1994; Hayes 1995). Thus, one can explain classical conditioning effects as caused by association formation in memory, or as the effect of pairing of stimuli. Negative priming can be explained as resulting from the inhibition of mental representations, or as resulting from the preceding presentation of a related stimulus, and so on.

It has been cogently argued that both these approaches have their merits, and that their combination may in fact be a source of progress for psychology (De Houwer 2014). However, the notion of causality may be a potential stumbling block. In particular, the evidence for causal claims in the cognitive approach is mainly probabilistic, since it deals with unobservable constructs, while the functional approach uses mechanistic evidence, since features in the environment can be observed or measured, and manipulated.

Does this mean that causal pluralism is the way to go? This would undermine the efforts of making the cognitive and functional approaches compatible. Moreover, in the context of the health sciences, Russo and Williamson (2007) argue that this move would confuse causal claims with the evidence for causal claims, and that instead an epistemic notion of causality is required that can account for the diversity of causal evidence – a situation that applies just as well in the case of cognitive psychology (or so I will argue). In this paper, I will argue that Menzies' functional-role account of causation is suited to this job. According to this account, causal relations are defined as playing certain functional roles with respect to abstract causal models, where a model is an ordered pair of the kind of system at hand, and the laws governing that system. In this vain, the causal relations between behavior and mental constructs, and between behavior and environmental factors, can co-exist as playing different, mutually compatible explanatory roles, and be argued for on the basis of diverse types of evidence. Thus, this account of causality has the promise to capture quite naturally the explanatory practice of both the cognitive and the functional approach, and hence to remove a potential obstacle to combining them.

Green, Sydney Katherine: A Defense of the Evidence-Based Medicine Model in Psychotherapy: Placebos and the Role of Culture in Healing
The appropriateness of the evidence-based medicine (EBM) model for research into the efficacy of psychotherapeutic techniques has been heavily debated. EBM, first developed within the context of pharmacology and surgery, holds that clinicians should practice only those therapies which have been shown to be effective in randomized controlled trials. It is argued that such studies provide the

only epistemologically-reliable means for evaluating a treatment's efficacy in remedying disorder, and for demonstrating a causal link between the treatment and the resulting cure. Therefore, if psychotherapy is to maintain the same rigorous scientific standards as pharmacology and surgery, it has no choice but to adhere to the evidence-based model. However, in direct opposition to this idea, many have argued that studying and evaluating psychotherapeutic techniques in such a manner is misguided. These critics argue that psychotherapy is fundamentally different from pharmacology and surgery, citing the disproportionately large role played by placebo effects in psychotherapeutic efficacy. They argue that the presence of placebo effects makes it difficult, if not impossible, to gather valuable information about causality in psychotherapy in a clinical trial setting. Thus, in place of EBM and the research methods it prescribes, they recommend various alternative approaches to studying psychotherapy, ranging from phenomenological studies of patient experiences to clinical narratives.

I argue that this rejection of EBM's use in psychotherapy is misguided. While others before me have attempted to defend EBM, they have not yet adequately addressed the specific criticisms of those who argue against its use. Rather than engaging with their critics, EBM's defenders too frequently give a blanket response to all criticisms, theoretical and practical alike: that there is simply nothing better out there. In contrast to these currently-existing defenses of EBM, my approach goes beyond this strictly negative argument. Instead, I use the arguments of EBM's critics to motivate its reconceptualization, not only in psychotherapy, but in other branches of medicine as well. I argue that the criticism of EBM is based on a flawed conceptualization of placebo effects. Taking inspiration from the anthropological work of Daniel Moerman, I argue that a more nuanced understanding of placebo effects, one which recognizes the causal role that culture and other social factors play in all healing processes – mental and physical alike – will show that there is no categorical difference between the way patients respond to pharmacological and surgical medicine on the one hand and the way they respond to psychotherapy on the other. Although this more nuanced understanding exposes certain methodological pitfalls present in psychotherapy trial design, I argue that it also offers ways of remediating those pitfalls, ultimately strengthening the epistemological power of the EBM model of research.

Grünbaum, Thor: Counting Computational Mechanisms

Milner and Goodale's theory of vision and action – the so-called Two Visual Streams Hypothesis (TVSH, for short) – has received a great deal of attention in cognitive neuroscience and philosophy of mind. TVSH plays a central role in philosophical discussions of the nature of perception and its content (Matthen 2005, Briscoe 2008, Wu 2014), the nature and function of consciousness (Clark 2001), the nature of action and control (Wallhagen 2007, Clark 2007, Mole 2008, Wu 2013) and the nature and status of folk-psychology (Bermudez 2006, Spencer 2007, Grünbaum 2012). For instance, Matthen (2005, e.g., p. 313) used

Milner and Goodale's TVSH to support the claim that visual object experience is characterized by a "feeling of presence". Clark (2001) has used TVSH to argue that visual consciousness is for high-level cognition and plays no role in "on-line" control of action. And the view has also been used by Bermudez (2006) and Spencer (2007) to argue for the falsity of our folk-psychological explanations of object-oriented actions in terms of visual experience of objects and conscious intentions.

These philosophical debates have been preoccupied predominantly with the issue of consciousness – issues such as the role of consciousness in control of action and the neural correlates of consciousness (e.g., Kozuch 2015). For the most part, they have been conducted as though the functional model of perception and action at the heart of TVSH (i.e. the functional perception-action model, or 'PAM' for short) had been uncontroversially specified. In fact, however, there is ongoing debate in the cognitive-psychological and neuroscientific literature about how best to specify PAM (Goodale 2014, Schenk and McIntosh 2010); moreover, it remains unclear how the specification of PAM relates to claims about consciousness. These ambiguities motivate the present paper. My aims are twofold: First and foremost, I identify a basic methodological problem that is inherent in studies designed to identify the correct version of PAM. Specifically, I demonstrate that scientific and philosophical discussions of PAM commonly assume a particular view of the individuation of computational mechanisms, and argue that if we assume this view of individuation, then behavioural evidence cannot be used to adjudicate between competing formulations of PAM. Secondly, if the argument of this paper is sound, then it has consequences for the use of Milner and Goodale's perception-action model to support strong claims about the architecture of the mind and the role of consciousness.

Kaiserman, Alex: Interventionism and Mental Surgery

The interventionist account of causation analyzes causal claims in terms of correlations under interventions, manipulations of variables which satisfy certain conditions. Campbell (2007) has recently argued that one of these conditions – the so-called 'surgical' constraint – runs into trouble when we try to use interventionism to model causation in psychology. The problem is that the constraint implies that an intervention on an intention should remove that intention from the influence of its rational causes. But it is implausible that our interest in psychological causation is an interest in what would happen in situations in which the rational autonomy of an agent is suspended in this way. "Someone who seemed to find him- or herself in that situation…would experience this as thought insertion, the feeling that someone else's token thought has been pushed into your mind…It is exactly this situation that we are envisaging, though, when we think in terms of surgical intervention on possession of an intention" (Campbell2007:62)

This paper defends two claims. First, I will argue that the problem Campbell raises for interventionism is in fact an instance of a wider problem, which I call the problem of 'abrupt transitions'. Suppose I'm driving in the right-hand-lane of a dual-carriageway, and realize too late, at time t, that I need to take the next exit. I miss the exit, and as such I am late for my meeting. If Campbell is right, an intervention on my being in the right-hand-lane at t would have to remove this event from the influence of my location at any earlier time. One can certainly imagine interventions like that – involving teleportation devices and the like – but it's implausible to suppose that our interest in whether my being in the right-hand-lane at t caused me to be late for my meeting is an interest in what would have happened had such bizarre conditions been in place.

David Lewis solves this problem by insisting that "we should sacrifice the independence of the immediate past to provide an orderly transition from actual past to counterfactual present and future" (Lewis 1979: 463). I show how to capture this solution within an interventionist framework. First, I show how the surgical constraint as stated by Woodward is ambiguous, and distinguish between a weak and strong version. I then argue that the problem of abrupt transitions arises only for the strong version. I conclude that the weaker constraint is the correct constraint to adopt. On this version of interventionism, an intervention on an intention to act can operate through the usual rational causes of that intention, so long as there is no independent causal route from the rational causes to the action. I illustrate this by discussing some actual interventions on intentions from the experimental literature (Webb and Sheeran 2006). I end by discussing familiar problem cases from the philosophy of action literature, arguing that my revised surgical constraint can capture exactly the difficulties raised by these cases.

Kästner, Lena Causal Mechanisms -- Not the One and Only ...

According to the mechanistic view, scientists explain phenomena by uncovering the mechanisms responsible for them (e.g. Machamer, Darden, and Craver 2000, Craver 2007, Bechtel and Abrahamsen 2005, Illari and Williamson 2012). What precisely this "being responsible for" means, however, is a matter of heated debate in contemporary philosophy of science (see e.g. Couch 2011, Harbecke 2010, Leuridan 2012). The suggestions vary significantly as an effect of what exactly we take to be "the phenomenon to be explained" and the "mechanism underlying it", respectively. Most prominently perhaps, we are told to conceive of the relation between mechanisms and their phenomena in causal or constitutive terms depending on their spatio-temporal characteristics and boundaries (see also Kaiser and Krickel forthcoming). While such metaphysical analyses certainly shed light on phenomenon-mechanism relations, they remain silent about the relations between different explanatory projects in empirical research.

In practice, mechanisms are supposed to explain a whole collection of different things: processes with multiple phases, end products or final stages of processes, stable states, continuous or repetitive behaviors, even properties. Naturally the

relationship between the "phenomenon" and the "mechanism responsible for it" will vary with the nature of the relata. A causal sequence can cause an end product to be generated, an arrangement of interactions can constitute an overall complex process, etc. While the explanatory projects focusing on causal and constitutive factors, respectively, are metaphysically quite different, scientists actually flip back and forth between them quite frequently even within a single research project. This is also highlighted by recent discussions on mechanism discovery. For instance, Craver and Darden (2013) suggest that researchers need to employ different strategies if they aim to find mechanisms that produce (cause), underlie (constitute), or maintain their phenomena, respectively.

While this may sound temptingly metaphysical, I suggest a somewhat pragmatic reading of the triade. Building on a well-known example from biology—lactose metabolism—I illustrate how scientists' search for explanations is guided by different research questions throughout the discovery processes. Depending on exactly what research questions they ask at any given point, scientists will differentially emphasize causal, constitutive, or continuous aspects in the mechanistic explanations they construct. There are two lessons to learn form this: (i) Causal explanations are undoubtedly an important part of the enterprise, but there is more. Indeed, insights about constituents or maintaining factors may be more important for some explanatory projects. And, (ii) to speak of mechanisms that produce, underlie and maintain their phenomena, respectively, does not mean that there actually are three different sets of goings-on in the world. Rather, it is a matter of perspective. What exactly we take "the phenomenon" to be depends on the explanatory project we pursue and which research questions we ask. What kind of mechanism can be "responsible for it" is determined by the nature of the phenomenon. This, in turn, determines not only the phenomenon-mechanism relation at hand but also whether our mechanistic explanations will emphasize causal, constitute, or continuous aspects.

Kleinberg, Samantha What causes a causal relationship?
While computational causal inference focuses on finding relationships between pairs of variables, many biological systems cannot be understood in this way. Our prior work on the physiology of patients recovering from stroke found that a loss of connectivity in the brain may lead to the emergence of new connections to maintain homeostasis. If we focus only on finding relationships between pairs of variables (whether this means edges in a graph, or quantifying the significance of each relationship), we would miss a key discovery here: that fewer causal relationships lead to the development of new ones. That is, it may not be the specific relationships that are important, but rather that a loss of connectivity leads to new connections, which may differ between individuals. This problem is tightly linked to that of nonstationarity, where causal relationships may change over time, and which has proven challenging for inference. In this paper we show how searching for causes of system changes can lead to better handling of nonstationarity and more informative inferences.

One of our key findings about stroke recovery is that whether a patient experiences brain swelling in part depends on brain oxygenation, but this relationship is modulated by the overall state of the patient. Swelling that develops earlier on during stroke recovery, in contrast, depends primarily on fluid status in the brain. Finding when and why such a relationship changes is key to effective treatment, ensuring that the actual cause of swelling is targeted directly. With traditional inference methods, though, such nonstationary time series are either transformed to be stationary (so this relationship will never be found) or involve partitioning the data into stationary subsets (which is computationally intensive and not always possible when features are continuously changing, such as during long-term weight loss).

More broadly, if the relationship between two variables changes (either emerges as a new connection, or changes in intensity) there must be a factor that modulates this. In the example above, patient state is a cause of the causal relationship between brain oxygenation and swelling. In general, however, this modulating factor may be an event (e.g. new law being passed) or a continuous change in a variable or property (e.g. weight loss over time). In both of these cases, it is important to note that it is not the event that directly brings about an effect, but rather the event modulates how other factors interact: creating a new connection, or strengthening/weakening existing ones.

In this paper we discuss how methods for causal inference from observational data can be adapted to handle this type of relationship, and illustrate our proposed approach using the example of stroke recovery. For infrequent, discrete, events we can use these event times to split the time series and then search for relationships within segments, instead of trying to find periods of stationarity. For continuously changing events, we show that we can recover functional relationships between a modulating factor and a causal relationship itself.

Kostic, Daniel Topological controllability of the brain: a case of non-causal interventionism

On Woodward's view, to explain an event or outcome is to provide "information about the causes of that outcome"(Woodward 2010: 291). Causes are understood along interventionist lines: two features are causally related just when, given some background circumstance, there is a possible intervention on the state of one feature that changes that of the other. The causal information is explanatory precisely because it can be used to answer what-if-things-had-been-different questions.

Topological approach also utilizes notion of interventions, in which features of network topology allow us to understand the system dynamics as a function of its structure. For example, the Watts and Strogatz (1998) small-world network model was built in such a way that starting from a ring lattice it has n nodes and k links. The structural properties of such a network are quantified by using its

characteristic path length, which measures a typical separation between two nodes in the network, which is expressed as L(p); and the clustering coefficient C(p), which measures the cliquishness of a typical neighbourhood of nodes. The small-world networks are characterized by low L(p) values, which is due to a few long-range links. Such 'short-paths' connect nodes that would otherwise be much farther apart and in effect shortening the path lengths between the whole neighbourhoods, and neighbourhoods of neighbourhoods. An explanation of why infectious disease will spread more rapidly though a population which instantiates a small-world topology, appeals to these structural features: pathogens can reach much larger number of nodes more rapidly if the L(p) is low and the C(p) is high. This explanation is specifically non-causal because the explanans cites only mathematical values of the model, not the causes.

Recent work in network neuroscience builds upon these ideas and conceives interventions through network control theory. It investigates how structural features of the brain networks determine temporal features of its cognitive dynamics (Bassett et al 2015). On this view neural system has a trajectory, which is a temporal path of the brain through various states, "...where a state is defined as the magnitude of neurophysiological activity across brain regions at a single time point." (Bassett et al, p. 2). Controllability is understood as a possible intervention on the mathematical measures of the structure that allows particular nodes (brain regions) that are at key locations in brain's topology to reach some other nodes. This measure allows us to find the brain region which is a control node, one that affects global topology, which ultimately affects temporal dynamics.

The significance of interventions is that they allow us to find explanatorily relevant variables. The only difference between the causal and non-causal interventions is what kind of explanatorily relevant variables they allow us to find. In the traditional interventionism it's causally relevant ones, in topological interventionism it's the properties of topology through which some activity drives the system to diverse states. Explanation of activity that drives the system and topology through which it's driven are distinct, the former is causal and the latter is non-causal, but both are interventionist.

Missal, Marcus: Experimental evidence in favor of a pluralistic conception of causality

Causal cognition encompasses different phenomena like causal inference and perception. Is the same causality concept valid in these different instances? In causal inference studies, causality is often conceptualized as a dependence relationship that can be intervened upon by operations on graphical models (see Danks 2014). In contrast, causality perception probably results from an inner representation of forces acting on objects. If two objects collide, we 'perceive' that something is transferred between them (energy, momentum) and we describe the resulting change of motion as causal. This causality 'perception' is congruent with the general definition of causality by transference: « Two events a

and b are causally related in the sense that one is a cause of the other if and only if there exists a conserved quantity Q of which a particular amount P is transmitted between a and b » (Kistler 1998). The two different conceptions of causality (transference and dependence) are very different and it could be suggested that in fact there is no single causality concept applicable in all domains of cognition. The aim of the present work is to further investigate this discrepancy using the approach of experimental philosophy. Two hypotheses could be distinguished. First, if there is somewhere in the human brain a single 'causality detector', as suggested by Albert Michotte (1946), then it is straightforward to suggest that its output could be shared in the different domains of causal cognition. Second, depending of the context, causal cognition could be based on transference or dependence. In order to test these hypotheses, we recorded eye movements and causality judgments of human observers watching a Michotte collision display. Indeed, when moving objects collide, a representation of physical causality in sensorimotor systems is necessary to guide behavior. Many characteristics of eye movements are not voluntarily controlled. Therefore, eye movements could be used to study the implicit representation of physical causality. We also asked the same observers to verbally state (answering 'yes' or 'no') whether a causal interaction between objects had taken place in the collision display they just watched and followed with the eyes. Surprisingly, we found that the same collision display yielded different responses: oculomotor behavior seems to be guided by physical causality (transference) whereas causality judgments varied on a subject-by-subject basis and as a function of the context (dependence). The higher-level causality concept subjects used to make a judgment did not overlap with the lower-level causality representation that guided behavior. In conclusion, if we accept that experimental psychology experiments could constrain philosophical investigations, our study supports the thesis of the pluralism of causality concepts.

Parkkinen, Veli-Pekka: Genetic causation in developmental and population behavioural genetics

This talk considers the concept of genetic causation in developmental and population behavioural genetics. The bulk of behavioural genetics research focuses on tracing genetic differences between individuals to differences in psychological phenotypes, providing clues of genetic origins of population wide variation in behavioural and psychological traits. By contrast, developmental behavioural genetics studies how genes dispose individuals to particular psychological phenotypes; the focus is on understanding causes of phenotypes, not phenotypic differences. In this talk it is argued that the population-level and developmental behavioural genetics – the latter understood as the study of determinants of individual phenotypes – employ different underlying concepts of genetic causation, and that confusion may arise when considerations and evidence pertinent to applying one concept are used in a context where the other concept is more appropriate. In the population perspective, genes are conceptualized as candidate actual difference-makers for phenotypes in a given environment: a gene may be said to cause a phenotype in the sense that allelic

differences map to phenotypic differences, and – per hypothesis – intervening to change the proportions of actually observed alleles in a population would change the distribution of the phenotypes. In the developmental perspective, genes cause phenotypes by determining the shape of the developmental landscape, making some phenotypes more accessible than others. The population perspective highlights genetic causes as loci of differences in a given environment, while the developmental perspective highlights "funneling" causation where genes have their influence through biasing development towards particular phenotypes against hypothetically varying environments. There seems to be a simple way of reconciling these perspectives – just think of the influence of genes on development as a mechanistically mediated effect, where genetic differences cause differences in developmentally intermediate physiological phenotypes that map to differences in psychological phenotypes that emerge later. However, this interpretation would ignore a potentially important role of genes in determining the sensitivity of development to environmental variation. Sometimes genetic differences result in phenotypic differences because an allele makes its bearer susceptible to particular environmental effects that, when the environment is present, push development to a path that is not as easily accessible for individuals not carrying the susceptibility allele. It is argued that in order to reconcile the individual and population perspectives, we should conceptualize development as a series of changes in the causal capacities of an organism over time, partly determined by its genetic makeup. Studies on the interaction of rearing environment and serotonin transporter gene promoter polymorphism in moderating depression are used as an example to illustrate these arguments.

Popa, Elena Understanding the Causal Asymmetry through Developmental Psychology

In this paper, I employ experimental data from psychology to connect the asymmetry of causation to the temporal asymmetry from the perspective of causal understanding. In philosophical context, asymmetry is a feature that makes causes have a certain priority over their effects (causes precede their effects and effects can be brought about through their causes, but not the other way around). In psychology this feature can be accounted for in terms of the distinction between causal and diagnostic reasoning (Fenker et al. 2005). My claim is based on recent data on the role of observation and intervention in children's learning of causal structures. The relevant experimental findings pertain to the use of temporal cues in causal learning tasks earlier in development, and to better performance in cases where intervention information is combined with temporal cues (Lagnado & Sloman 2004, McCormack et al. 2014,), I argue that, as far as causal understanding is concerned, the asymmetry of causation can be explained through the temporal asymmetry. On my view, causal understanding follows the developmental pattern where causal relations are in first instance understood through their asymmetric feature, and thus as subject to temporal direction, and later on difference-making information comes in (given that not all

claims about temporal sequence result in causal claims - spurious causation cases, for instance). This claim is supported by the fact that in experimental setting whenever interventions are performed, they are subject to a temporal delay (Lagnado & Sloman 2004).
From a more general perspective, the two types of evidence for causal claims in psychology under investigation fit the framework of Hall's (2004) distinction between causation as production and causation as difference-making. The former, involving a spatio-temporal process from cause to effect can accommodate temporal evidence as relevant for causal claims, while the latter emphasizes the role of difference-makers (for my purposes, intervention data). I hold that both kinds of evidence can be accommodated, with temporal evidence being understood earlier on. In broader philosophical context, this claim draws a different perspective on the relation between causal understanding and temporal direction. While authors such as Price and Weslake (2009) argue that both temporal direction and the causal asymmetry can be understood through the deliberation situation, my claim is that, rather, the causal asymmetry could be understood through one of the features of the deliberation situation: the fact that all the decisions and acts of the agent are subject to temporal direction. This way, both temporal and intervention evidence for causal claims can be accounted for, with an explanation of people's preference for temporal cues when inferring causally. Furthermore, the reliance on psychological data could deal with one of the problems of equating the causal and temporal asymmetries: simultaneous causation cases. Experimental data (Bramley et al. 2014) shows that people find it more difficult to build their causal claims on simultaneous causation cases. On my view, while simultaneous causation is metaphysically possible, it is more difficult to grasp by causal reasoners.

Sprenger, Jan: Foundations for a Probabilistic Theory of Causal Strength
Causality is usually treated as a qualitative, all-or-nothing concept. However, causal inference in science often requires a quantitative dimension, that is, judgments of causal strength (or causal effect). For instance, regression coefficients in linear models and synaptic weights in connectionist networks can naturally be interpreted as quantifying the strength of a causal relation (Fitelson and Hitchcock, 2011). In the latter case, the concept of causal strength can be used for quantifying the effect of an experimental intervention, or for attributing cognitive abilities to the activity of specific neurons.

The paper assumes a probabilistic account of causal relevance (causes raise the probability of the effects), amended with the manipulability view of causation and Pearl's do-calculus for describing interventions. That is, C is a cause of E if and only if an intervention on C (written do(C)) would change the probability that E takes a certain value. Causal strength is then determined by a directed acyclical graph (DAG) which describes causal dependencies between the

relevant variables, and a joint probability distribution of these variables (for forerunners, see Good, 1961; Cheng, 1997). This approach does not only square well with recent theories of causality (Pearl, 2000; Woodward, 2012), but also with statistical measures of effect size, and research on the intersection of probabilistic and causal reasoning (e.g., Sloman and Lagnado, 2015).

The paper proceeds as follows. First, I outline a framework in which different measures of causal strength can be defined and compared. Second, I show that no single measure can capture all intuitive demands on a concept of causal strength. Third, I prove two representation theorems for natural and much-discussed measures of causal strength. That is, I show how the measures in question can be derived from a set of plausible adequacy conditions. Below, I state the main results of the paper in an informal way.

Theorem 1 We need a plurality of measures of causal strength. (No measure that is sensitive to probability-raising can satisfy our intuitions about actual causation.)

Theorem 2 The Pearl-Suppes measure $\eta_*(C, E) = p(E|do(C))$ is an adequate measure for describing the degree to which an observed effect can be attributed to a cause.

Theorem 3 The Eells measure $\eta_d(C, E) = p(E|do(C)) - p(E|do(\neg C))$ is an adequate measure for quantifying the predictive value of an intervention and the size of an observed effect.

The paper is primarily descriptive, but to the extent that the adequacy conditions are plausible, it has normative implications for the choice of a measure of causal strength. It is also methodologically innovative in transferring methods from probabilistic reasoning, confirmation theory and theoretical psychology to a quantitative model of causal reasoning (Cheng, 1997; Crupi et al., 2013).

Based on these results, I develop a taxonomy for probabilistic measures of causal strength, argue why the two above measures should be treated as defaults, and dis- cuss an application to the mind and brain sciences: quantifying causal strength in medical psychology. In particular, I demonstrate why the measure of Absolute Risk Reduction (ARR), which corresponds to the Eells measure η_d, should be preferred to its competitors as a measure of risk reduction for categorical variables (cf. Stegenga, 2015).

Wallot, Sebastian: Interaction-dominant vs. component-dominant causation in mind and brain, and their implications for questions of generalization and replication.

In this paper I want to discuss two views on causation in the sciences that investigate human behavior and neurophysiology. According to Van Orden (2003), the majority of research in those sciences subscribes to the view that the

dynamics of behavioral and neurophysiological measurements are dominated by a finite set of time-invariant components to which observed variance can be attributed. For example, behavioral and neurophysiological measurements during reading of this abstract can in parts be attributed to distinct memory components (long-term, working, perceptual...) and brain regions/networks. Hence, explaining observed dynamics in brain and behavior means to identify and chart the different components involved. The notion of causation implied here is that mind and brain work in a way so that some input for a participant in an experiment is always mediated by the same components in the same (qualitative) manner to link this input to some measurement – simple feeback loops notwithstanding. The causal chain is hard-assembled. In contrast, from the view of interaction-dominant dynamics such component-causal chains are soft-assembled. Evidence for such soft-assembled, interacition-dominant causation comes from the pervasiveness of statistical interaction effects (Abraham, 1987) or complexity characteristics (Jensen, 1998) in measures of brain and behavior: For example, when memory functions and brain regions/networks change in their number or function with changes of the input (e.g., reading a random word list instead of this abstract), then this produces a statistical interaction effect, whereby variance in the observed measures cannot be attributed to a specific component (if one assumes that word and paragraphreading are at least somehow comparable activities). In the case of complexity characteristics, such as the presence of ‚turbulence' in measures of brain and behavior, the problem is that variance that ought to come from a specific mental or neurophysiological component cannot be separated from noise characteristics of the observed measure. In effect, it becomes difficult or impossible to isolate component effects in the dynamics of an observed measure. In so far as research in the sciences of mind and brain is confronted with continued interaction effects or complexity characteristics, one can assume that a modeled causal chain of compoments is only soft-assembled, and that any explanation based on this model resides above, and not at the level of the actual, invariant causal architecture that governs dynamics of mind and brain. Moreover, this has implications for quesitons of generalization and replication: Scientists cannot – by default – assume generalization of a causal model to a new measurment situation. Also, interaction dominant dynamics might be an alternative explanation to the current replication-crisis in the behavioral sciences, as some of the failures to replicate reported effects is not necessarily only due to flawed research and data-analysis, but due to an wrong assumption of stability of causal component-chains, namely that they are by default stable over arbitrary changes in time and location.

Wilde, Michael: Extrapolation and mechanisms

In the context of animal modelling, extrapolation often involves attempting to determine whether a causal claim established in a model organism is also established about the target organism. For example, it may be established that a particular exposure causes cancer in rat models, and the problem is determining

whether the exposure also causes cancer in humans. This problem is often confronted by the International Agency for Research on Cancer (IARC). The IARC *Monographs* aim to identify exposures that cause cancer in humans by evaluating a range of evidence, including evidence from both studies of cancer in humans and studies in experimental animals.

It has been popular to appeal to mechanisms in helping to solve the problem of extrapolation (see, e.g., Schaffner (2001) and Weber (2005)). The basic premise is that extrapolating a particular causal claim from an animal model to humans is reliable only to the extent that the animal model is analogous to humans in terms of the relevant mechanisms (see, e.g., Weber (2005: 154-7)).

Daniel Steel (2008) has objected to this mechanistic approach on the grounds that it fails to address the challenge of the extrapolator's circle. In particular, if the evidence of shared mechanisms is sufficient to establish both the causal claim in the animal model and that the animal is a reliable model for extrapolation to humans, then presumably this evidence is sufficient also to establish the causal claim in humans without the need for extrapolation. This is because establishing the causal claim in the animal model would be consistent with the corresponding causal claim not being established in humans, unless the evidence of shared mechanisms also established the causal claim in humans.

Steel proposes comparative process tracing as a solution to the extrapolator's circle. This involves first learning the mechanism in the animal model and then comparing the relevant mechanisms in humans, at key stages where the mechanisms are likely to be different. Steel suggests that the greater the similarity of mechanisms at these stages, the stronger the basis for extrapolation.

I argue that a couple of case studies in the IARC *Monographs* provide some support for the claim that comparative process tracing helps to determine the reliability of an animal model for extrapolation. Aristolochic acid was evaluated as carcinogenic to humans on the basis of having identified key stages in the mechanisms by which tumours had formed in experimental animals and humans as a result of exposure to aristolochic acid. For instance, stages in the mechanisms by which aristolochic acid causes tumours in experimental animals had been identified, and these mechanisms were consistent with events occurring in patients with urothelial cancers associated with aristolochic acid nephropathy. The particular mutations identified in humans with aristolochic acid nephropathy were present also in experimental animals exposed to aristolochic acid. I argue that this looks a lot like comparative process tracing, and in this case it confirmed the reliability of the animal model for extrapolation. (I also consider another case study involving the evaluation of d-Limonene.)