Interventionism, Realism and Invariance

=

The kind of metaphysics that matter

Mori Diakite

Master's thesis in Philosophy

Department of Philosophy, Classics, History of Art and Ideas
Faculty of Humanities

Supervisor: Researcher Anders Strand

UNIVERSITY OF OSLO

Spring 2016

1 Preface

Preface: a short characterization of the sum of reflections, affections and experiences that constitutes the student's journey from the first sentence in the introduction to the last sentence in the conclusion.
How I would characterize and summarize my philosophical journey?

As a journey filled with trials and tribulations that taught me a valuable life lesson.
Namely that, a necessary and sufficient condition for reaching my goals, is that my determination remains invariant under transformation of challenges.

I would like to give a special thanks to each of my family members, Emerencé, my mother, Olga, my aunt, Adam, my brother, Aminata, my sister, for always believing in me. - The invariance of your support caused my determination to remain fixed.
I would like to my gratitude to Farzad, my teacher, and each of my friends and loved ones, that despite the absence of my presence, continued to cheer for me till the end.

Finally, I would also like to give a special thanks to both of my supervisors Anders Strand and Gry Oftedal for being supportive and giving valuable guidance and help. I would also like to extend this to my former supervisor when writing my bachelor's thesis, Veli-Pekka Parkinnen.
If it was not for you, I might never have come across the Woodward way.

2 Summary

The title of the thesis "Interventionism, realism, and invariance – the kind of metaphysics that matter" gives away two of the central topics that drives the thesis, (1) the intersection between interventionism, realism and invariance, (2) that the subject matter is concerned with metaphysics. Interventionism, refers to James Woodward's account of causation that centers around the idea that causal relationships are relationships that are exploitable for purposes of manipulation and control.

The manipulation and control under question refers to a special kind of manipulation, namely, an intervention, a surgical process that for heuristic purposes, might be thought of as a controlled human experiment.
The interventionist aims at capturing the two following principles:

(DM) The Principle of Difference-Making: for X to be a cause of Y, is for X to make a difference to the conditions of how Y obtain or is altered. More precisely, if the relation between X and Y is causal, then this relation qualifies as difference-making if X were to obtain, that would make a difference to the conditions of how/whether Y obtains or is altered.

(CR) The Principle of Change-Relating: For X to be a cause of Y, is for X and Y to be linked by a causal process such that, if changes in Y occurs, they do so in virtue of changes occurring in X. More precisely, provided that the relation between X and Y is causal, this relation is change-relating in the sense that if X were to change this would affect the manner in which Y changes.

In order for the two above principles to be successfully captured within an interventionist theory of causation is for the following condition to hold: *The relationship between X and Y must remain stable or unchanged as various other changes occur***. – That is, the relationship must exhibit invariance under intervention**, which allows us to correctly assess and capture the correlation between the two variables under question**.**

In the thesis I argue that the commitment to realism, the idea that there exist objective, mind-independent causal relationships (and facts about the truth of causal claims) conjoined with the above conception of causation, interventionism, and the related necessary and sufficient condition for a relationship to be causal, invariance under intervention, introduces metaphysical commitments.

In light of the idea that it is built into the interventionist conception of causation that there exist objective, mind-independent causal relationships (that are exploitable for purposes of manipulation and control), I argue that metaphysical commitments are built into, and follows from an interventionist theory of causation (insofar as includes realist commitments, as Woodward does in (Woodward 2003, page 121)).

3 Table of Contents

4
Introduction: Interventionism, realism and invariance; the kind of metaphysics that matters

Woodward opens up the chapter on Invariance in Making Things Happen by noting the following; “the guiding idea is that invariance is the key feature a relationship must possess if it is to count as causal or explanatory. Intuitively, an invariant relationship remains stable or unchanged as various other changes occur” (Woodward 2003, page 239)

This thesis discusses the notion of invariance under intervention, with particular emphasis on its role within the interventionist framework, and the metaphysical implications that surrounds it.
The first aim of the discussion of invariance is to highlight what kind of metaphysical commitments Woodward himself admits, while in turn theorize about the nature and scope of these metaphysical commitments.
The second aim will be to examine whether these admittedly metaphysical commitments are sufficient to substantiate the role that invariance is supposed to fulfil, or whether Woodward should commit himself further in order to cash out his theory.

A common theme running through Woodward’s exposition of the metaphysics of interventionism and invariance under intervention, is that the notion of metaphysics that figures in the interventionist framework is deflationary[1], which translates to metaphysical

[1] It should be noted that there is a dynamical development in the stance that Woodward takes towards the status, role and position of metaphysics within the interventionist framework. – Woodward’s position changes from taking a quietist approach to metaphysics in the form of dismissing the need to engage with it (considering that he regards it as a separate (and partly irrelevant) project in contrast to what he is interested in (as he describes it in Woodward 2003; investigate the semantics of causal claims). This stage is marked by his staunch criticism of the idea that every account of causation (and explanation) requires an account of the truth-makers of causal claims (as is highlighted quite explicitly in the exchange between Woodward (2007) and Strevens (2008)).
In later writings such as Woodward 2007, we notice that Woodward abandons the quietist approach by engaging critically with questions and topics that concerns the metaphysics of causation and explanation, such as causal foundationalism and the truth-conditions for causal claims in science.
However, it is in the latest writings such as Woodward 2015, that development of a position and thesis about metaphysics takes form.

commitments being limited to what he terms 'modest realism', the idea that there are determinate, mind-independent, objective (facts about) truth-values of counterfactual conditionals concerning what would happen under combination of interventions (Woodward 2003, page 121)– This attitude towards metaphysics follows from the methodological underpinnings of interventionism, that leads to a functionalist approach to both causation (see Woodward 2012) and metaphysics (see Woodward 2015).

My thesis counteracts this 'deflationary' approach to metaphysics by arguing for two propositions; under the heading of my central claim in the thesis:

Central claim in thesis: Invariance under intervention introduces metaphysical presuppositions and implications as a consequence of its role within the interventionist framework and the logic of interventionism. – These extend beyond Woodward's characterization of 'moderate realism'.

(1) Proposition: *That metaphysical commitments are built into interventionism*, particularly invariance under intervention due to the logic of interventionism (logic refers to the methodological and ontological considerations that are part of the application, assessment and structure of interventionism).
(2) Proposition: that the metaphysical commitments of interventionism, particularly invariance under intervention, *extends beyond the deflationary attitude and restrictions that Woodward admits to*, which leads to a commitment to 'substantial' realism, the idea that metaphysical commitments, in addition to requiring objective facts about counterfactuals, requires further commitments concerning the structure of causal relationships (particularly their mind-independence and objectiveness).

Overview of (some) of the central arguments for my thesis, and the corresponding two propositions:

Woodward builds upon Hitchock's development of the metaphysical thesis and framework, "means-ends metaphysics", that develops a conception of metaphysics that integrates methodological and pragmatic considerations into the framework (the means), while constraining the metaphysical input/output of the framework in terms of the same considerations that are present in the framework (ends). – Woodward extrapolates and builds upon this framework as to develop his own account of means-ends metaphysics framed in terms of the interventionism. – The central idea is that in virtue of distinguishing between different conceptions of metaphysics and different goals for the inquiry of metaphysics, then interventionism can be relevant for metaphysical inquiry as a contributor to methodology (how we inquire about metaphysics).

The claim: Invariance under intervention, serving as the 'mark of causation' and pillar of the logic of interventionism, introduces metaphysical presuppositions and implications. → Metaphysics is both built into and follows from the logic of interventionism.

Justification for holding this claim: The conjunction of causal correctness, (counterfactual) modality and realism/objectivity, carries metaphysical commitments and has metaphysical implications due to the role and commitments of invariance under intervention.

Elaboration of the justification I: I contend that, metaphysics as understood in an interventionist framework must be construed in a gradual manner, in which the extent of metaphysical commitments is a function of the range, scope and depth of invariance.

Elaboration of the justification II: I propose that the whole idea of invariance under intervention presupposes an ontological distinction between boundary/initial conditions and invariance conditions, where the former is open to variance, and the latter exhibits invariance. – The ontological basis for the distinction follows from the idea that we locate invariances in 'the world' rather than (or at least not restrictedly to) 'our models'[2].

Elaboration of the justification III: I contend that invariance (modularity) within our representational devices and models, insofar as the representation is supposed to be veridical (correctly representing the causal system under investigation), requires an account of how modularity relates to the system under investigation, such that we can read off invariance from modularity (as when we describe correlations among a set of data) or vice versa, (as when we predict correlations on the basis of observations) and hence conclude that there is a causal relationship on the basis of this reading. – In line with the realist underpinnings of interventionism, I further claim that this relation (between modularity and invariance) will be asymmetric (the invariances in the world determines the modularity within our models, but not vice versa)

Remark concerning the relation between functionalism and metaphysics:
As a way of making the transition from Woodward's functionalist framework to a metaphysical inquiry, I do not want to pause at the question of whether Woodward's theory provides an account of truth-makers for causal claims (as authors such as Strevens have

[2] As a consequence of realist underpinnings.

opted to do), but rather I want to assess how and to what extent the logic of interventionism incorporates metaphysical commitments when characterizing causal relationships. – This remark explains the considerable emphasis on the methodology of interventionism.

As an illustration of some of the theoretical background that I draw upon when proposing the above analysis of the metaphysics of invariance under intervention, I refer to Cartwright and Nozick that has both proposed similar treatments of the logic of invariance.

Cartwright in her paper "Two Theorems on Invariance and Causality" notes a similar conclusion (albeit slightly different terminological and definitional understanding of both interventions and invariance), by drawing a connection between the three terms, interventions, invariance and causal claims (Cartwright 2003, page 204). – In this paper, she pursues the following goal: "Then, given some natural and relatively uncontroversial assumptions, I prove two distinct sets of theorems showing that invariance is a mark of causality when the concepts are appropriately interpreted" (Cartwright 2003, page 204). The focus is not on Cartwright's conclusion and arguments, but rather on the conditions she puts forward for discovering and assessing the relationship between invariance, causation and interventions (that I both agree with and will pursue in this thesis):
"To get clear about whether invariance under intervention is or is not necessary or sufficient for a causal-law claim to be correct, and under what conditions, we need to know what counts as an intervention, what invariance is, and what it is for a causal-law claim to be correct." – Cartwright 2003, page 204)

Robert Nozick in his work "Invariances: the structure of the objective world" provides an analysis of invariance under admissible transformations, that despite its differing subject matter (concerns the relation between invariance and objectivity with a basis in physics, rather than invariance and causation as in Woodward's framework) shares a great deal of similarities when it comes to understanding the logic of invariance. – The central feature that is common between the two, is the idea that invariance is a partial and relative notion in the sense that there will be a domain of invariance that, depending on the set of admissible transformations (transformations are restricted to interventions in Woodward's framework, whereas Nozick leaves the transformations undefined for generality), will allow us to distinguish between the degree, scope and characteristics of invariance depending on

the metrics we utilize for assessing it, and the dimensions (background conditions) we assess it along and against. – I will return to Nozick's framework in the chapter on invariance.

With this introduction in place, I turn to characterizing the methodology of interventionism, as that will set the precedence for the rest of the discussions that will follow.

5 When methodology meets ontology; a functionalist approach to causation

Does an account of causation require a conceptual analysis that aims at reducing it to non-causal terms in order to be illuminating? If the focus is ontological (rather than conceptual), does the explanatory prowess of an account of causation rest on its ability reduce causal relations to non-causal correlations? Does nomothetic laws play an indispensable role in every successful causal explanation? Does the commitment to the objectivity of causation and explanation presuppose that accounts of causal explanation avoid pragmatic considerations as they threaten to undermine objectivity[3]?

James Woodward with the help of his interventionist account of causation and causal explanation (henceforth interventionism for brevity) answers these questions negatively. Interventionism is non-reductive (in the sense of analyzing causal claims in terms of causal notions), denies the traditional assumption that causal explanation requires invocation of laws (in the sense of proposing an alternative approach to causal explanation and revising the notion of laws that figures within it), while instead of dismissing the status and role of pragmatic considerations, integrates them within the interventionist framework.

Woodward develops a framework for understanding the semantics of causal claims, the relation between causation and causal explanation, and the role of laws in terms of interventions.

I will get into detail of some of the central features of this framework in the next chapter, but in this chapter I want to focus on the methodological considerations that figures in the interventionist account of causation.

I want to discuss three topics in this chapter, (i) why methodology, and if so, how? (ii) What

[3] The idea that is expressed by the contrastive focus between objectivity and normativity is meant to express the contrast between objectivity as a mind-independent feature that depends on the world and normativity as a mind-dependent feature that depends on our interests. – If the prerequisite for establishing and maintaining the objectivity of causation rests on the idea that an account of causation should omit references to non-objective features, then normative and pragmatic features, in virtue of being mind-dependent, will threaten to undermine the objectivity of the account under investigation.

is the exact relation between methodology and ontology? (iii) How does methodology relate to my thesis?

As a preliminary background for discussing the source of methodological (and pragmatic) considerations, we might pose the following question: why care about methodology in the first place?

A suitable point of departure is discussing the guiding idea behind an interventionist notion of causation, namely that "causal relationships are relationships that are exploitable for purposes of manipulation and control".

Woodward notes on the first page of 'Making Things Happen' that "an interest in causes and explanation pervades our lives", which serves as an illustration that causation and explanation are something that are of interest to us as agents.
As an attempt to explain the origin of our interest in causes and explanations, Woodward proposes the following explanation: "(…) Our interest in causal relationships and explanations initially grows out of a highly practical interest human beings have in manipulation and control: it is then extended to contexts in which manipulation is no longer a practical possibility" (Woodward 2003, page 10).

In order to expand this explanation (especially the part that concerns practical possibility) while combining it with the modality (a feature that figures extensively in the characterization of interventions, and the interventionist framework as a whole), Woodward notes that (...) our interest in causal explanation represents a sort of generalization or extensions of our interest in manipulation and control from cases in which manipulations is possible to cases in which it is not, but in which we nonetheless retain a concern with what would or might happen to the outcome being explained if various possible changes were to occur in the factors cited in the explanans." (Woodward 2003, page 11).

What follows from these considerations, is that human agency and practical interests (in manipulating the world) play an important role in illuminating our notion of causation and explanation, which in turn illustrates how methodology enters the interventionist picture. Methodology will be a dimension of interventionism that concerns; (i) the conditions under which our account(s) causation and explanation maps into our practical interests, (ii) the pragmatic features that figures in our accounts of causation, (iii) the commitments/implications of admitting methodological considerations within an account of

causation and explanation. – These are the features that I'm dedicating the current chapter to illuminating, with particular emphasis on (iii).

As an illustration of the impact of agency and practical interests within the interventionist framework, I refer to what Woodward himself has to say about why and how these considerations are illuminating.
As a reply to the question of what agency and practical interests might be informative and important, Woodward answers; "these interests explain or largely explain why we have a notion of causality at all, and why it takes the form or has the features that it does", and further, "(…) enables us to understand why causal claims have many of the features they do and helps to adjudicate between rival claims about those features" (Woodward 2003, page 150-151).
These two examples are just a selection of the importance that is tied to pragmatic and methodological considerations, it remains to clarify the full-scale extent, weight and importance of these considerations, something that I know turn to providing a glimpse of.

To hint at my conclusion, the status, role and implications of methodological and pragmatic considerations turns interventionism into a functionalist account of causation. This influences the status and role of metaphysics within the interventionist framework, in the sense that methodological considerations both constrains and guides the nature of metaphysical commitments and implications. – Woodward (2015) elaborates upon these constrains and guiding principles and develops a coherent notion of metaphysics, namely, means-ends metaphysics.

Before turning to discussing the details of the methodology of interventionism, I conclude the preceding section by noting that the answer to the question we started at the beginning of this chapter, why care about methodology in the first place?
Is answered by the central role that methodological and pragmatic considerations play within an interventionist framework. – The source of methodological considerations follows from the guiding idea underlying an interventionist conception of causation, namely that 'causal relationships are exploitable for purposes of manipulation and control'.

5.1 4.1 Methodology as a pragmatic enterprise – activities, interests, functions; the beginning of a functionalist account of causation

Woodward (2003) provides a preliminary account of the methodological consideration that underlies his interventionist framework.
Starting with the manipulationist intuition (that causal relationships are exploitable for purposes of manipulation and control) and the observation that our interests in causation and explanation derives from our interests as practical agents, Woodward goes on to propose a methodology for causal explanation (includes causation as well)[4].

The subject matter of these methodological considerations are not restricted to the means, strategies and purposes of causal inquiry and explanation, but concerns metaphysical considerations such as the role /requirement of reductionism, and epistemological considerations, such as the epistemic constraints on causal information.

Woodward's preliminary account of the methodology of causal explanation can be divided into four tenets; (1) that causal explanation is a practical activity, (2) that his account of causation and explanation is non-reductive, (3) that there are epistemic constraints on causal explanation, and (4) that there exists a set of desiderata for an account of causal explanation. I will restrict my discussion to (1) and (4) as they are the most relevant for highlighting the methodological underpinnings of interventionism (whereas (2) and (3) represent consequences of these methodological underpinnings).
Due to space considerations, I will aim at presenting the broad lines in these methodological tenets in order to highlight their contribution to substantiating the connection between the manipulationist intuition and methodology of interventionism.

5.1.1 First tenet; causal explanation as a practical activity – making the case for the connection between causation, explanation and action.

This tenet follows from the manipulationist intuition and have implications for the methodology, practicality and epistemology of interventionism.

[4] It should be noted for future references that during the discussion of the methodological considerations that figures In Woodward (2003), the use the term 'causal explanation' includes references to causation as well.

As hinted at earlier, Woodward conceives our interest in causation and explanation as an extension of our role as practical agents that act with the intention of realizing certain goals (Woodward 2003, page 18). If we unpack this further, the idea seems to be that practical agents, in virtue of specifying a goal, being aware of some selection of courses for action, and attaining knowledge of the consequences that follows from them [the actions], will select a course of action that produces the consequences that best matches their specified goal.
The process of specifying goals, selecting actions, and in turn, relating their consequences to their compatibility with goals, maps nicely into the manipulationst intuition as our interests will figure as a point of departure for causal inquiry, while the compatibility between our goals (purpose of manipulation and control) and selection/consequences of actions (information about how facts about dependency relations relate to purposes of manipulation and control), will be guiding principles for assessing causal relationships.

Insofar as we start with a specified goal and select a course of action, we are able to learn about its consequences, and by learning about the consequences, we are able to revise and/or select a course of action that best matches with our goals. – In other words, there will be a feedback loop between selection and assessment.

Relating this to causation, this process yields grounds for distinguishing causal relationships from non-causal correlations (as the former is useful for purposes of manipulation and control, while the latter is not).
Furthermore, our account of causation will be guided by the compatibility between our goals (purposes of manipulation and control) and selection/consequences of features (information about how facts about dependency relations relate to purposes of manipulation and control), which in turn guides which features we include in our account of causation (as Woodward himself claims in; Woodward 2003, page 149).
The account of causal explanation that follows from this logic produces knowledge of causal relationships (which starts with a clear connection to practicality and agency, but extends to circumstances under which agency and practicality is irrelevant).
This is the first consequence of considering causal explanation as a practical activity.

Methodology enters the picture through describing the circumstances, conditions and requirements under which our interests, assessment and actions mesh. Epistemology enters the picture through imposing a requirement that we have epistemic access to the information that figures in the assessment of causal relationships (as I will illustrate in a moment).

Practicality enters the picture insofar as the manipulationist intuition figures within the interventionist framework.

As for the second consequence of considering causal explanation as a practical activity, I can be more brief when highlighting the connection between the manipulationist intuition and methodology of interventionism.

Woodward starts from the (uncontroversial) assumption that most (if not all) communities and/or societies have and occasionally do engage in explanatory activity (read as including reference to causal inquiry). Furthermore, Woodward recognizes the distinction between the explanatory activity that goes on in everyday/pre-scientific circumstances that all human societies/communities engage in, and the more systematic and sophisticated activity that goes on within contemporary scientific communities. – As Woodward states quite explicitly, by recognizing causal explanation as a practical activity, we should expect a continuity between these two senses of explanatory activity on both a 'substantive' and 'methodological' level (Woodward 2003, page 19).
The substantive level concerns and refers to the content of the explanations, whereas the methodological level concerns and refers to the goals, interests and structure that underlies the explanations.
On a substantive level, the continuity takes the form of the scientific causal explanation integrating and developing the knowledge of prior causal relationships that we utilize in everyday context. On a methodological level, the continuity explains how different scientific causal explanations shares some of the structural features that are found in causal explanations formulated in ordinary contexts (pre-scientific). – Distinguishing and referring to the two levels as 'substantive' and 'methodological', might give the impression that the former level is more fundamental and/or important than the latter, but in reality, this terminology is only meant to demarcate the different aspects of the continuity that they relate to.

The continuity does not only let us encompass causal explanation (in a general sense) within a single framework, but it also lets us generalize the manipulationist intuition to every context where explanatory activity and causal inquiry is present.
The continuity is a consequence of the generality of the manipulationist intuition that in Woodward's sense is meant to encompass causal inquiry and explanatory activity in general. The differences between the forms of explanation is explained in terms of their appeal to

knowledge with varying degree of rigor, depth and detail.
As to how this aspect of the interventionist framework connects methodology and manipulation, this connection is reveled in the way that causal inquiry and explanatory activity requires a continuity between both content (substance) and form (method). The fact that this continuity is a presupposition, and that it is integral to the formulation of our accounts of causal explanation, illustrates how manipulation and methodology is invariably connected to each other.

5.1.2 Desiderata: what we should expect from a theory of causal explanation? – making the case for balancing between description, prescription and revision

This section of Woodward's methodology might be divided into three parts; (1) the nature of his project (compared to traditional approaches), (2) the components of the desiderata that makes up his methodology (and how they are intertwined), and (3) general prescription/desiderata for a theory of causal explanation. – I contend that it is the conjunction of the desiderata and manipulationist intuition that marks the clearest emphasis on methodological considerations in the interventionist framework.

In Woodward (2003), interventionism is described as a project that balance between the two traditional philosophical projects of conceptual analysis (description of "cause" and its related locutions and usages) and metaphysical analysis (description of the 'nature' of causation and causal relations). In Woodward (2012), interventionism is described as relating to three projects (with interventionism representing a fourth project); (i) metaphysical project, (ii) descriptive project, (iii) 'fit with physics project', and (iv) functionalist project[5].

In light of the fact that Woodward's the backbone of interventionism is the manipulationist intuition, and that part of Woodward's strategy and methodology is to extend this intuition to the interventionist framework as a whole, we might present the desiderata as a set of expectations that connects the different aspects of interventionism to the manipulationist intuition.

[5] My focus will be on describing the methodological considerations that figures within the formulation of interventionism as a functionalist project (Woodward 2012) as it draws broader lines to topics I will discuss later (metaphysics + methodology versus ontology). However, considering that Woodward (2003) lays the groundwork and goes into much more detail on the methodological aspect, I will use the account in (2003) as a basis, while use (2012) to draw the broad lines and make the transition to the next topic.

In addition to balancing between the two above projects (conceptual and metaphysical analysis), integrating the practices of causal inference (and explanation) to the subject matter of conceptual analysis, Woodward introduces two additional components to the methodology of interventionism; 1. Intentionality (purposes and goals of our practices of making causal/explanatory inferences), and 2. Normativity (revisions and recommendation about what we ought to mean when making causal/explanatory claims) (Woodward 2003, page 7). – Commenting on the relationship between pursuing the activities of description and revision of our notions of causation and explanation, Woodward contends that "(…) both should be pursued together" (Woodward 2003, page 8).

Putting these observations and considerations together yields the following list of theoretical aims of the desiderata for his account of causation and theory of causal explanation:

1. Description of the usage of causal terms and practices of causal inference/explanation,
2. Prescription of the practices and usages in terms of their ability to realize the goals we set for them,
3. Revision of the meaning of causal notions and claims in light of what we ought to mean when using them (normative considerations).

As mentioned earlier, Woodward extends the domain of conceptual analysis beyond its traditional borders (traditionally restricted to analysis of meaning and usage of causal notions), by not only including an analysis of the practices of causal inference (and explanation), but also including a revisionary aspect of conceptual analysis (rather than settling for description of the meaning and usage of causal terms). When it comes to Woodward's relation to the 'metaphysical' analysis, his approach distinguishes itself from the traditional approach by relating the discussion of causation in to both normativity and practicality in addition to ontology, while the traditional approaches has tended to restrict themselves to the latter.

Extending these theoretical aims, and in turn extrapolating their domain of application, translates into general requirements for the content, structure, and aims of our accounts and theories of causation and (causal) explanation. - The consistent theme is the dual emphasis on description and revision with the aim of developing a coherent theoretical framework.
In light of the assumption that "we introduce concepts and characterize them in certain ways, partly because we want to do things with them", our accounts and theories should address the

content and utility of those concepts, and finally, assess them in light of whether they succeed in fulfilling the aims of their employment (Woodward 2003, page 8)

First, in order for anything to qualify as an account of causation (explanation etc.), whether descriptive or prescriptive, it must be constrained by prior usage and practices (Woodward 2003, page 9). The prior usage and practices provides a continuity of methodology and development between the old accounts and the new ones, while at the same time demarcates the subject matter of the account under study.
The aim of introducing this requirement, is that we (at least according to the interventionist) formulate accounts with the aim of fulfilling certain goals. Prior usage figures as a metric for both assessing and guaranteeing that we in fact (or at least get closer to) succeed(ing) in fulfilling these goals.

Second, in order for an analysis to increase the clarity of the meaning of causal terms and notions, it must include a revisionary aspect (in addition to descriptive) in the analysis of how ordinary folks (or experts) use causal terms such as "cause" and "explanation" (ibid).
The revisionary aspects provide a tool for addressing, and in turn correcting the errors and mistakes we make when making causal claims and using causal notions.
The aim of introducing this requirement is the underlying idea that our theorizing and analyzing is goal-oriented in the sense that our usage and assessment of the above terms and notions is influenced by the purposes of introducing them (remember the feedback loop between selection, assessment and goals from the discussion of 'practical activity').
As a way of getting closer to, and in turn, realizing the goals that we set out for our analyses and usage of causal terms and notions, we introduce the tool of revision and the idea of a feedback loop (between goals, assessment and selection) to provide a theoretical basis for this process.

This is a summary of the above discussion of the methodological, substantial and epistemological consideration that make up the methodology of interventionism (as formulated in Woodward 2003).
Starting with the manipulationist intuition (that causal relationships are exploitable for purposes of manipulation and control), and the observation that our interests in causation and explanation derives from our interests as practical agents, leads to the development of a methodology that groups Woodward's account of causation, causal explanation and causal inference into one framework that allows us to assess their content in an accessible manner in

terms of their ability to fulfill the purposes that we employ them for.
With this information in place, I want to turn to the implications of the above discussion, a functionalist account of causation.

5.2 Interventionism as a functionalist account of causation – interventionism, methodology and usefulness

In the preceding section, we provided an account of the source, content and aim of interventionism, with particular emphasis on its methodology.
The above account was based on the detailed explication Woodward (2003) provides of the methodological considerations that figures within his interventionist framework with particular focus on how the manipulationist intuition influences the different aspects of the interventionist framework (causation, causal explanation, causal inference).
However, despite the detailed exposition of the methodological considerations that figures in the interventionist framework, I only commented briefly on the implications of relying extensively on methodological considerations and how this relates to other topics such as ontology and epistemology.
in this section, I want to draw attention to one of the central implications of centering methodological considerations at the heart of interventionism, namely that, interventionism is transformed into a functionalist approach to causation.

As a way of characterizing what a 'functionalist approach to causation' consists of, and how interventionism is an instance of this approach, I will rely on Woodward (2012) paper that characterizes its broad lines and implications, while making references to Woodward (2003) in those situations where the broad lines fail to convey the details of its logic.

In the previous section we described the methodology of Woodward (2003) as a balance between the two traditional projects, conceptual analysis and metaphysical analysis, while making references to the developments and revisions that Woodward includes in his balanced and revised form of conceptual analysis (including revision and normative assessment of both causal terms, notions inference and explanation).
In Woodward (2012), interventionism is described as a lone-standing project that, instead of balancing between different projects, represents an independent project that share some of the

features with its competing projects, but stands in contrast with all of them in some way or another (Woodward 2012, page 3).

As we noted in the previous section, Woodward (2012) describes interventionism as relating to three projects (with interventionism representing a fourth project); (i) metaphysical project[6], (ii) descriptive project[7], (iii) 'fit with physics project[8]', and (iv) functionalist project. Of the mentioned projects, the functionalist project stands in stark contrast with the metaphysical project (due to the latter's heavy focus on metaphysics and 'metaphysical foundations' (as Strevens (2007) would characterize it), while the former insists on remaining non-committal when it comes to metaphysics and foundations). – I will restrict my attention to addressing the relationship between the functionalist project and the metaphysical project as my interest is in the relation between methodology and metaphysics (and not methodology in general).

Despite already characterizing and touching upon the central features underlying the functionalist project (goal-orientation, interest-relativity, and practicality) in Woodward (2003), we might list up the features of the functionalist project as Woodward (2012) describes them (albeit in a brief fashion due to space considerations).

The functionalist project takes as its point of departure: "the idea that causal information and reasoning is sometimes useful or functional in the sense of serving various goals and purposes that we have" (Woodward 2012, page 3). – Woodward describes the functionalist conception of causal cognition as a form of "epistemic engineering", drawing focus towards the

[6] As Woodward characterize this project: "Those pursuing this project think it important to provide a metaphysics for causal claims or to specify what causation "is", metaphysically speaking, or what the metaphysical "truth-makers" or "ground" for causal claims are." (Woodward 2012, page 1) – Examples/proponents of this kind of project: Tooley (1977), Armstrong (1983), Bird (2005).
The central feature that underlies these projects is their extensive reliance and emphasis on metaphysics in their discussions of causation.

[7] Similarily, as Woodward characterize this project: "Those pursuing the descriptive project attach considerable importance to constructing accounts whose aim is to describe or reproduce (what they take to be) the causal judgments of "ordinary folk"." (Woodward 2012, page 2) – Examples/proponents of this kind of project: Lewis (1973), Schaffer (2000), Hall (2004).
The central feature that underlies these projects, is in addition to describing the intuitions and judgments of "ordinary conception" of causation, to characterize them in terms of their ability to overcome hurdles/paradoxes such as; cases of overdetermination, preemption, transitivity, and temporal asymmetry (to name the most famous ones).

[8] Finally, as Woodward characterizes this project (whose name characterizes both its subject matter and enterprise): «focuses on issues having to do with the relationship between causal claims, including the sorts of claims made in ordinary life and in the so-called special sciences, and what is imagined by some philosophers to be «fundamental physics» (Woodward 2012, page 2) – Examples/proponents of this kind of project: Field (2003), Loewer (2009), Frisch (2014)

functionalist ideas of engineering, where the causal cognition functions as a piece of "epistemic technology", a tool that is assessed in terms of its ability to serve our goals and purposes.
This point of departure leads to the following three conseuqnces for the conception of causation:

1. Illustrates the importance of connecting causal notions to other concepts that we employ (including reasoning patterns and procedures for investigating nature). (ibid). – This consequence follows form the idea that causal cognition and the idea of usefulness (of causal thinking) admits of a plurality of strategies for causal cognition that is relative to context, purposes and interests.
2. Focus on methodology, or to be more precise, "on normative assessment (and not just description) of various patterns of causal reasoning, of the usefulness of different causal concepts, and of procedures for relating causal claims to evidence". (ibid)
3. The commitment to non-reductioninism. – Provided that a functionalist characterization and analysis of causal cognition and thinking is possible in the absence of a reduction to non-causal categories, there seems to be no reason to require reductionism (on functional grounds). (ibid)
4. Besides these, there are additional consequences such as the assessment of the various testing procedures for causal claims, the utility of the various distinctions we employ in causal thinking, and other normative questions and consideration (these are not relevant when it comes highlighting the central points of this section).

An important divergence between the methodology and goal-orientation in Woodward (2003) and Woodward 2012), is that the latter adds various other goals to the list of 'goals or purposes of causal thinking' (as we might remember, the former focused solely on the manipulationist intuition, the purpose and goal of manipulation and control, when addressing the question of goals within the interventionist framework).
Some of the goals include: "compact and unified representation of relationships useful for prediction", and, "the codification of our commitments to various inductive strategies" (Woodward 2012, page 5). – The focus is not on which goals that are added (or subtracted), but rather the extension of the functionalist ideas that underlies interventionism, which expands the framework, such that it might include other goals, other purposes, and hence, other strategies.

An important expansion of the functionalist framework in Woodard (2012), is its addressment of the metaphysical project and metaphysics as a whole. – What I want to draw attention to, is the way that functionalist ideas both constrains and guides the metaphysical commitments that we include into our accounts of causation.
Woodward (2012), restricts the metaphysical commitmetns of the functionalist framework to so-called 'modest realism'.
Modest realism consists of the following two claims (much more will be said about this):

1. That there is an objective difference between merely correlational and causal correlation.
   In other words, this translates to the fact that it is the world (and not just our interests) that determines, fixes and demarcates causal (from non-causal) relationships (Woodward 2012, page 7).
2. That causal relationships have an objective basis. In other words, the idea that it is the world (and not just our interests) that determines which causal relationships holds and in what circumstances (ibid).

The difference between the functionalist and metaphysical project, is that the former constrains metaphysical commitments by relating it to the purposes and aims of causal thinking (which in practice translates to refrainment from stipulating further metaphysical commitments unless they serve a specific purpose and that there is a clear aim of remaining non-committal when it comes to metaphysics), whereas the latter 'fails' to tell a story of how metaphysics relates to causal thinking (Woodward 2012, page 7). – As will be discussed in detail, the relation between functionalism ad metaphysics plays an important role in the interventionist framework, and is further developed by Woodward (2015) under the heading of 'means-ends metaphysics'.

A further difference between these two projects, relates to their diverging positions when it comes to the issue of causal claims and truth-makers. An uncontroversial and non-committal formulation of what truth-makers 'are' is this: if there is some causal relationship R that holds, which on an interventionist reading translates to, there being some intervention on X that is associated with changes Y, then there will be some deeper explanation of for why R holds. Both the functionalist and metaphysical project embodies this formulation, but they diverge in their analysis of what the content of these truth-makers are/would be.
From the perspective of Woodward's functionalist project, the content of the truth-makers,

will consist of "ordinary empirical facts and ordinary causal/nomological relationships (…)", while the deeper explanation of why R holds will consist of "(…) and the sorts of explanations they provide for claims like R will be ordinary scientific explanations." (Woodward 2012 page 8). As from the perspective of the metaphysical project, the content of the truth-makers will consist of the same empirical facts, relationships and explanations, but there will be an additional story of the metaphysical 'facts', 'entities', 'relationships', 'categories', or '[insert favored metaphysical term here]' that underlies the empirical facts under question.

Woodward concedes to the proponents of the metaphysical project that causal claims have, and is in need of truth-makers (viewed in light of Woodward's general quietist approach to metaphysics, such as in MTH, makes this a considerable concession), but denies the claim that these truth-makers need an additional [metaphysical] story besides the ordinary empirical facts. – The central point I want to draw attention to is not Woodward's general position towards truth-makers, but rather his reason for denying the requirement of the additional metaphysical story. As Woodward puts it "(…) I don't see why this sort of interpretation is required [referring to the additional metaphysical account of the empirical facts], at least for functional purposes" (Woodward 2012, page 8).

This is another illustrative example of how functional purposes constrains and guides the metaphysical considerations that figures within the interventionist framework.

As to summarize the current discussion, we started out with investigating the methodological considerations that underlies interventionism as formulated in Woodward 2003. This provided us with some loose threads that pointed in the direction of functionalism, the thesis that our interests and purposes constrains and guides our causal thinking and cognition. In Woodward 2012, we connected those loose threads, characterized them in terms of a unified framework, and finally drew out some of it's implications. The two central implications that have been the focus so far has been (i) methodology (what it is, why it is, and how it is) and (ii) ontology (what it is, why it is, how it is). Up untill recently, the main focus has been one methodology, with brief comments on the status and role of ontology (the comments have merely served as illustrations of how methodology might constrains and guide ontology).

From here on and out, my aim is to focus on ontology with the aim of answering two questions; what role does ontology serve within interventionism? Provided that ontology is interest-independent (modest realism, objectivity and mind-independence), how does this relate to the interest-dependent interventionist framework?

5.3 Methodology meets ontology – implications of conjoining functionalism and realism

In the preceding section we alluded to tying the methodological pieces of interventionism together, while formulating the implications of the methodological orientation of interventionism, namely that, interventionism embodies a functionalist account of causation. At the end of the previous section, I commented briefly on some of the implications of relating the functionalist underpinning of interventionism to metaphysics.
In a summarized fashion, given functionalism, metaphysical commitments are both constrained and guided by the purposes and interests of our causal thinking (however, given a modest realism about metaphysics, the world, and not our interests alone will influence the methodological underpinnings of a functionalist account of causation). (Woodward 2012, page 3, 7-8).

In this section I want to expand the discussion of the relation between methodology and ontology, particularly, the implications of conjoining a functionalist account of causation with a realist conception of metaphysics.

In order to motivate the study of this relation and provide some background for the discussion, I will present some of the central points of disagreement between Woodward (2008) and Strevens (2007, 2008), as a large chunk of their disagreements results from the former pursuing a methodology-first strategy and the latter pursuing a metaphysics-first strategy in the study of causation.

5.3.1 Strevens versus Woodward – semantics, ontology and metaphysical foundations

Strevens (2007) in his review of Making Things Happen (2003) after having sketched Woodward's causal manipulationism (in a more metaphysical guise and terminology than Woodward himself does in MTH and would sit comfortable with) discusses Woodward's account of event explanation and theory of type-level causation, before returning to criticizing the latter (Strevens 2007, page 233). – It is when discussing the theory of type-level causal claims that Strevens' metaphysical orientation appears in its complete guise, and it is this section of his review that I will center my presentation of his views around.

As a preliminary remark, Strevens (2007) defines causal manipulationism (Woodward's theory of causation) as "(…) the doctrine that the facts about causation metaphysically depend in part on the facts about what can be manipulated by what" (Strevens 2007, page 234).
According to Strevens' characterization of Woodward's views, type level causal relations enters the picture by providing "the metaphysical basis of causal explanation, by determining the facts about singular explanation and causal relevance" (Strevens 2007, page 244).
Strevens' centers his criticism around the issue of truth-conditions, where according to him, causal manipulationism fails to realize Woodward's aim of providing an account of "the content or meaning of statements of type level relations (…)", due to there being reason to doubt whether his theory fits the description (of the aim) if the content of such statements are supposed to fix their truth-conditions (ibid).

Part of what Strevens characterizes as fixing the truth-conditions of statements and in turn provide a metaphysics for them (in a causal manipulationst framework): formulating a 'non-relativized' notion of causation that operates with a non-circular definition of interventions that grounds the causal claims under question (either in non-causal correlations or some fundamental non-interventional notion) (Strevens 2007, page 244-245).
Relativization refers to Woodward's utilization of the idea of direct causation that is relativized to a variable set, and where Strevens' argues that a non-relativized notion of causation is required for capturing the "content or meaning of our causal claims" (Strevens 2007, page 244).
According to Strevens, that leaves us with two choices when assessing whether Woodward's relativized notion of causation fits the ticket; either (1) it figures as 'the fundamental causal notion' (which according to Strevens is at odds with 'our notion of causation' that is non-relativized) or that, (2) it helps us define 'a non-relativized notion of causation' that captures the "content or meaning" of our causal claims (Strevens 2007, page 244).

In conclusion, on Strevens' view, the kind of project Woodward is engaged in (providing an account of the content or meaning of type level causal claims) requires a metaphysical, ontology-first approach.
As to Strevens' verdict in light of this requirement; "the manipulationist account of type level causal relations is, I think, a strategic error (…)", which leads to the further conclusion that causal manipulationism "(…) is compatible with the rejection of metaphysical manipulationism, which is a matter not just of entailment but also of ontological dependence

[of facts about causation and facts about manipulation]" (Strevens 2007, page 247). Drawing out the consequences of the above assessment, Strevens notes: "Woodward's type level metaphysics should be abandoned, along with the relativization of causation", as a consequence of its failure to provide 'metaphysical foundations' for type level causal relationships (ibid).

Strevens ends his review by noting the valuable contributions of MTH in illustrating "the scientific role of the connection between causation and manipulation", but limits his praise in a way that highlights both the source of his criticism and the alleged deficiency of causal manipulationism; "(…) But do not make the mistake of thinking that its importance is reason to philosophically apotheosize manipulation as the foundation of everything – or perhaps anything" (Strevens 2007, page 249).

5.3.2 Woodward versus Strevens – pragmatics, methodology and anti-foundations

In his reply to Strevens, Woodward (2008) begins with pointing out that Strevens' basis of criticism and attribution of positions rests on misinterpretations of MTH and that there are two sources that both causes and explains the disagreement among the two; (1) "One reason for this is that Strevens reads me with the preoccupations of a metaphysician, (2) "another is that he relies heavily, in reporting what he takes to be my views, on restatements of those views within his own terminology and system of concepts rather than on what I actually say" (Woodward 2008, page 193).

Strevens' presents and characterizes MTH as, (among other things), an attempt to provide a metaphysics of causation (with the following descriptions and phrases of Woodward's project illustrating this point; "facts about causation metaphysically depend on what can be manipulated by what,", "Direct causation is the fundamental causal notion"). Woodward argues that it is this (wrongfully) metaphysical characterization that drives Strevens' assessment of MTH, as his criticism of MTH rests "(…) on the grounds that it is efficient qua metaphysics and lacks adequate metaphysical foundations" (Woodward 2008, page 193).

In an attempt to reconstruct Woodward's reply to this supposed mischaracterization, I want to focus on answering two questions: 1. What kind of philosophical enterprise does MTH (and interventionism as a whole) engage in, and how does this enterprise relate to the topic of

metaphysics? 2. What exactly is MTH, if not a 'metaphysical treatise' aiming at establishing 'metaphysical foundations' for causal relationships?

The answer to the first question, consists of a number of parts, each making up what Woodward takes to be the thrust of both MTH/interventionism and the philosophical enterprise it engages in, namely, methodology.
The primary focus of MTH is methodological (although it ranges over a number of topics), and its domain of application is philosophy of science, or to be more precise, the philosophy of causation with particular emphasis on "(…) how we think about, learn about, and reason with various causal notions and about their role in causal explanation, both as these occur in common sense and in various areas of science" (Woodward 2008, page 104). – In answering how his conception of the philosophical enterprise of MTH relates to Strevens' more metaphysically oriented foundationalist conception, Woodward notes that "the overall perspective of MTH is what might be described as that of a modeler-pragmatic, piece-meal, and anti-foundational" (Woodward 2008, page 195).

The implications of this 'overall perspective', is that we get a contrast between two approaches to the philosophy of causation, the one represented by Strevens which we might name an ontology-first approach, and the one represented by Woodward, by which we might name a methodology-first approach.
Both of them agrees that an analysis of causation includes investigating the semantics of causation and causal claims, which includes the usage of 'definitions' and 'truth-conditions' to capture "the content and meaning of causal claims and notions" (Strevens 2007, page 244) However, the great rift between the ontology-first and methodology-first approach occurs during the explication of the relation between truth-conditions, definitions and the aims of causal inquiry.

The proponent of the ontology-first approach would like to say something like; the semantics of causation, insofar as it concerns truth-conditions, is first and foremost cashed out in terms of capturing the dependency relations in 'causal reality'. – This is another way of stating that semantics should be pursued by doing metaphysics, as the former is built into the latter (at least this seem to be the position that Strevens (2007) advocate).
The proponent of the methodology-first approach disagrees and responds with something like: the semantics of causation, regardless of being concerned with truth-conditions, can be illuminative even though it does not cash it out (only) in terms of capturing the dependency

relations in ‘causal reality’ (where ‘capturing’ is understood as a form of reduction, as will be highlighted later). – This is another way of stating that there is no implicit connection between investigating semantics and doing metaphysics (as Woodward (2008) seems to be of the opinion that it is possible to investigate the one without the other)

Relating the methodology-first approach characterized in Woodward (2008) to the question of how MTH relates to metaphysics, seems to give the impression that MTH does not relate to metaphysics at all (if so, it is in a limited sense or as an offshoot of the other considerations that figures within the interventionist framework) in light of the fact that the interventionist project is

As to the answer of the second question concerning the character of MTH, Woodward proposes that one of the attractions of the manipulationist account of causation (the on figuring in MTH) is its “unmetaphysical character” that opens up the possibility of thinking of causation in terms of manipulation and control (rather than traditionally metaphysical ideas such as relations of necessitation among universals, similarity relations among possible worlds, just to name a few possibilities) (Woodward 2008, page 194). – This illustrates what MTH is not, namely a metaphysical treatise arming at providing a metaphysics for causation.

As to illustrating what MTH is, has been alluded to earlier, which is a functionalist account of causation (as we have discussed it in detail earlier, there is no need to reiterate the same points, therefore I’ll rather restrict myself to highlighting the further implications for the assessment of ontology). – From this ‘unmetaphysical’ functionalism follows the fact that definitions in MTH are to be understood ‘instrumentally’, namely as tools that are ‘(…) to be judged by their usefulness for various purposes (rather than if they capture fundamental metaphysical relationships), and where examples of metrics for assessing this usefulness are ‘exhibiting connections and interrelations with other causal notions, ‘connecting with issues about information about causal relationships that are learned, and ‘contrast with other treatments of causation in the philosophical and scientific literature (Woodward 2008, page 195-196).

The central implication of MTH’s unmetaphysical methodological approach that is characterized in Woodward (2008), is the two-fold sketch of a methodological approach to metaphysics that consists of the following to theses; (i) distinction between causal inquiry and metaphysical inquiry (with the two being separate and independent of each other, such that it

is possible to investigate the one without the other), (ii) utility of causal inquiry despite the absence of 'metaphysical significance' (with the basis of utility following from pragmatic/methodological considerations that does not presuppose nor imply metaphysics) (Woodward 2008, page 197). – In contrast to the position towards metaphysical commitments in MTH (2003), with its modest realism (that aims at limiting the role and centrality of metaphysics in the interventionist framework), the position of Woodward (2008) towards metaphysics seems to have shifted towards a methodological a-metaphysical position (that aims at demarcating metaphysical from non-metaphysical inquiry, while at the same time restricting interventionism to the latter form of inquiry).

However, despite this maneveur, Strevens (2008) in his reply to Woodward (2008) notes that the a-metaphysical position mounted by Woodward (2008) in defence of his MTH (2003) is at odds with a number of theoretical goals that is pursued in MTH.

Strevens draws attention to three examples from MTH that states that contrary to the alleged a-metaphysical and methodological demarcation of interventionism from metaphysics, MTH addresses and engages with metaphysical issues (Strevens 2008, page 182). As Strevens puts it; "What I cannot give you is a non-metaphysical goal for a philosophical account of causation that is a plausible interpretation of what Woodward is up to in *Making Things Happen* – because despite his denials, what he is up to is pretty clearly metaphysics." (ibid). As to the examples; (i) the contrastive analysis of his account of causation and (clearly metaphysically oriented) accounts of Lewis's counterfactual accounts and Menzies and Price's manipulation account that groups them together in the same 'genre' (Lewis 1983, 1986b, 2000; Menzies and Price 1993). (ii) the references to features of causation that are incorporated into the interventionist account of causation, and that are clearly metaphysical in nature such as the idea that causation is 'mind-independent' and 'objective' (Woodward 2003, page 118). (iii) The commitment to truth-conditional analysis of causation that follows from, and is incorporated into the semantics of causation, which according to Strevens is thought of "(…) as aiming to specify those representations' truthmakers. It may look like semantics, but it is also a kind of metaphysics".

Despite these clear examples of aspects of MTH that engages with metaphysics (either directly or indirectly), there are passages in MTH that reiterates the deflationary attitude towards metaphysics; "Beyond this [the modest and noncommittal formulation of realism found in MTH], it commits us to no particular metaphysical picture of the "truth-makers"."

(Woodward 2003, page 121).

This leaves us at a crossroad between the metaphysically modest formulation of interventionism, that limits and restricts the amount of metaphysical commitments, and the a-metaphysical formulation of interventionism, that demarcates and distinguishes causal inquiry from metaphysical inquiry. The Common thread running through both of these formulations is the emphasis on methodology, I was planning to unpack this further in the next section by going into detail on Woodward (2015) formulation of metaphysics, means-ends metaphysics, a methodology-first, deflationary, yet realist approach to metaphysics. But, unfortunately due to lack of space and sufficient amounts of times I did not get the opportunity to do so.

6 Connecting the dots: when interventions and causation meets invariance conditions

The aim and function of this chapter is to provide the needed background conditions for discussing the content, form and implication of invariance under intervention. According to Woodward, Invariance, which is short for invariance under intervention, is the distinguishing feature that causal claims and causal relationships needs to exhibit if they are to qualify as causal and explanatory (Woodward 2003, page 239).

What I want to do, is restricting the discussion to the relationship between interventions and invariance conditions, as the main thrust of invariance, is the way it incorporates interventions into its characterization/description of clarifying what, how, and why we should integrate invariance conditions into our accounts of causation, laws, and explanations.

The connection between invariance and explanation is not integral to the characterization of invariance under intervention, but it proves to be a central way of highlighting the applicability of invariance. – Therefore, as a way of getting to the heart of invariance conditions and their metaphysical implications, I will restrict the focus to interventions, invariance, and relatedly, causal generalizations as they figure in causal relationships and explanations. I will mainly mention or refer to causal explanation as a way of highlighting the applicability of invariance under intervention.

The examination of the relationship between invariance and interventionism divides into two parts: (i) elaborating on what the interventionist account of causation and explanation consist of, and (ii) highlighting how and to what extent this account presupposes invariance. This two-fold examination will not be restricted to clarifying the relationship between interventionism and invariance, but will also aim at highlighting and disentangling the complex interplay between pragmatics, metaphysics and methodology that characterizes Woodward's treatment of the above themes.

6.1 Interventions, causal assessment and invariance – An interventionist theory of causation

In the preceding chapter we provided a preliminary account of what interventions consists of and the role there are meant to fulfill in the assessment of causal query. The picture we drew of interventions depicted them as hypothetical/ideal experiments that serves function and role of causal assessment, namely, ideal experiments that determines whether and in what sense a correlation between two variables, X and Y, qualifies as causal (rather than mere correlational). When relating interventions to causal relationships, we did so with the intention of capturing and elaborating on the two intuitions that the interventionist analysis assumes, namely, difference making (DM), the idea that a cause makes a difference to its effect, and, change-relating (CM), the idea that the changes in the cause leads to changes in the effect.

The unpacked and more detailed formulation of these ideas (for future references):
(DM) for X to be a cause of Y, is for X to make a difference to the conditions of how Y obtain or is altered. More precisely, if the relation between X and Y is causal, then this relation is difference-making in the sense that if X were to obtain, that would make a difference to the conditions of how/whether Y obtains or is altered.
(CR) For X to be a cause of Y, is for X and Y to be linked by a causal process such that, if changes in Y occurs, they do so in virtue of changes occurring in X. More precisely, provided that the relation between X and Y is causal, this relation is change-relating in the sense that if X were to change this would affect the manner in which Y changes.

In this section, we will elaborate on how these ideas are captured within Woodward's interventionst framework, in particular, how the semantics of interventionism encapsulates the ideas of difference-making, change-relating and interventions. In doing so, I will distinguish between two aspects of the semantics of interventionism. (i) the aspect that is concerned with the conditions and structure of causal relationships, and (ii) the aspect that is concerned with the details of the relationship between interventions and causal relationships, notably, the relationship between interventions, variables and the fixation of values (of variables) in the causal relationship.
To draw parallels; this distinction corresponds to the distinction that Woodward makes between two types of information, the information about variables that has been set to some value by an intervention, and information about the causal systems that the variables are embedded in and which remains intact despite intervening on one of its variables. (Woodward 2003, page 47)

6.2 Interventions, causal structure and causal relevance – highlighting what we mean by causation

Returning to the preceding characterization of interventions and the ideas of change-relating and difference-making, what we are interested in, is highlighting how these translate into conditions that needs to be in place for the claim X causes Y to come out as true (understood along interventionist lines). The question is thereby, what are the truth conditions for causal claims understood along interventionist lines? Moreover, what is the exact connection between manipulations/interventions and causation?
The beginning of an answer to the preceding questions follows naturally from examining the conditions that the interventionist theory of causation introduces to highlight the connection between interventions and causation. The connection between interventions and causation takes the form of a question that leads to an answer that consists of two conditions; a sufficient and necessary condition for a relationship to be causal (formulated along interventionist lines.

Sufficient condition (SC): "If (i) there is a possible intervention that changes the value of X such that (ii) carrying out this intervention (and no other interventions) will change the value of Y, or the probability distribution of Y, then X causes Y." (Woodward 2003, page 45)

Necessary condition (NC): "If X causes Y then (i) there is a possible intervention that changes the value of X such that (ii) if this intervention (and no other interventions) were carried out, the value of Y (or the probability of some value of Y) would change." (Woodward 2003, page 45)

These conditions exemplifies how interventions are linked to causal relationships, they are processes that if carried out/realized changes the value of the cause (X) such that the value of the effect (Y) changes because of the intervention on the cause. – This gives us the following truth condition for a causal claim along interventionist; for X causes Y to be true is for there to exist an intervention with these properties (the ability to change the value of Y by changing the value of X).
As was discussed briefly in the previous chapter, interventions should be understood hypothetically, which does not require the invocation of human agency and physical possibility (this is respected in the preceding conditions by talking about "possible interventions" and inserting "if… then" clauses). – In light of these remarks, we should

interpret the invocation of interventions existentially, which is another way of stating that all that is required of the truth conditions to be satisfied, is that it is possible (arguably not impossible) that there exist an intervention that stand in a relationship to the causal claim under question.

For future references, I will refer to the commitment to the existence of an intervention (satisfying the specified conditions of changing the value of Y by changing the value of X) as:

The commitment to the existence of possible interventions: "The truth of the claim (understood along interventionist lines), X causes Y, commits to the existence of a possible intervention, I, that changes the value (or the probability distribution) of X, which then changes the value (or the probability distribution) of Y". – Woodward 2003, page 45

However, regardless of these clarifications, there are a number of pressing questions that lurks in the background, for instance, what is it to "change the value of" of the variables in the causal claim? What exactly is the causal relata in such causal claims, and how does it influence the interpretation of such claims? What kind of possibility does interventions require and how are we to make sense of it without invoking physical and/or practical possibility?
Each of these questions will be addressed in turn, but what I want to draw attention to is the idea that causation is linked to the change of values, as that allows us to discuss both the change-relating aspect of the causation and the causal relata within the interventionist framework simultaneously.
To clarify the relation between change-relating (the changing of values) and causal relata, the natural point of departure is to examine the conception of 'change' that Woodward is interested in.

Definition of what it is to "change" the value of a variable: "If we think of an intervention as the taking of a particular value by an intervention variable I, then I changes the value of X and, in doing so, changes the value of Y if and only if there are values x_1, x_2 of X with $x_1 \neq x_2$ such that I's assuming some value i_1 causes X to assume the value of x_1 and I's assuming some distinct value i_2 causes X to assume the value of x_2 and there are values y_1, y_2 of Y associated with x_1 and x_2 such that $y_1 \neq y_2$. In other words, if $Y = F(X)$, describes the functional relationship between X and Y, there are distinct values of X, x_1 ,and x_2, and values of Y y_1, and y_2 such that $y_1 = F(x_1) \neq y_2 = F(x_1)$. – (Woodward 2003, page 45)

As was hinted at earlier, and as Woodward states in this excerpt, on an interventionist account of causation, causal relationships relates variables, or to be more precise, relates changes in the value of one or more variables with changes in the value of other variables (Woodward 2003, page 39). Unpacking the idea of change a bit further yields the following proposal; when a variable, V, is changing, it takes different values, v, along a continuum $v_1, v_2, \ldots\ v_n$, such that, for an arbitrary variable V_i to undergo change, is for the variable to take a different value $V_i \ldots V_n$,where $V_i \neq V_n$. The last equation that states that the value of V_i and the new value V_n, is different (not identical) can be generalized into a requirement that each of the values along the continuum is discrete, which is another way of stating that for every value $v_{i,} v_n\ of\ V, v_i\ \neq v_n$. The discreteness of both variables and values of variables will prove to be indispensable when it comes to making sense of combining talk of variables assuming different values, with talk of combining different interventions to assess the causal relationship between variables.[9] More will be said about this in subsequent sections, notably, when talking about the characteristics of variables and interventions as causal relata.
Furthermore, and this is the crucial step, the changes we are interested in assessing are those that results from interventions, particularly, those that follows from the intervention variable I, assuming a value I_i that is connected to the value of a variable V_i. The crucial changes are those that occur because of I_i assuming an arbitrary value I_n, which then causes the variable V_i to assume the value V_n. In this way, we can map the value of the intervention variable to the value of the variable we are assessing, which in turn gives us the following functional relationship: $I_1 = V_1, I_2 = V_2, \ldots\ I_n =$
V_n $(condition\ of\ discreteness\ applies\ here\ are\ well)$. It should however be noted that there is a restriction on the sort of changes in the variable, V that count as interventions, namely the changes "satisfy whatever conditions must be met in an ideal experiment designed to determine whether X causes Y", which reinforces the earlier remarks about interventions serving the role of causal assessment (Woodward 2003, page 46).
If there is an association between V_i and a second variable Y_i, such that whenever $V_1 =$ $Y_1, V_2 =\ Y_2 \ldots\ V_n = V_n$, provided that the changes of the value of V_n can be attributed solely to

[9] This is especially important when it comes to variables that are embedded into a larger model that contain several other variables (that may or may not have a causal influence on the relationships we are assessing), which requires more fine-grained information about the dependency relations among the variables. Woodward expresses a similar condition when requiring that an interventionist analysis of causal claims presupposes that there is a "well-defined" notion of changing the value of a variable, which do not only concern the relation among the different values that a variable might take, but which also concern the very idea of applying an interventionist analysis of causal claims. (Woodward 2003, p. 111, p. 113)

the changes of the value of I_n, then the same goes for the value of Y_n, such that we have a causal link running from $I_i \rightarrow V_i \rightarrow Y_i$. However, for this last inference to succeed in assessing the nature of the association, in other words, whether V causes Y, it is necessary that the changes that are caused by the intervention satisfy a second condition; arrow-breaking.

The idea of arrow-breaking in the present context is the following.
"More generally and slightly more precisely, we may think of an intervention on X with respect to Y as an exogenous causal process that changes X in such a way and under conditions such that if any change occurs in Y, it occurs only in virtue of Y's relationship to X and not in any other way". (Woodward 2003, page 47).

Unpacking further, the idea is that if we have a causal structure that consists of several variables, and that these variables are associated in the sense of exhibiting influence on each other, then we have a structure that represents endogenous causal relationships. What an intervention is meant to accomplish, is "breaking" this endogenous causal relationships, in other words, interventions are supposed to be exogenous changes in the variable intervened on, such that it disrupts the influence exhibited by the other variables, on the variable being intervened on, such that the value of the variable is fixed entirely by the intervention.

For future references, I'll label this requirement as:
The commitment to interventions qualifying as exogenous and arrow-breaking processes: interventions are exogenous processes that changes the variable intervened on by altering the causal structure that the variable is embedded into. This alteration disrupts the causal influence(s) previously exhibited by other variables, on the variable that is intervened on, such that the value of the variable is fixed entirely and solely by the intervention (while leaving the other facets of the causal structure intact).

This concludes the detour we took from the discussion of the conditions SC and NC, to clarify and illustrate the connection between change-relating and variables (and their values). To sum up, change understood along interventionist lines is understood as variables taking different values. The changes that are central to the assessment of causal relationships (and claims) are those that occur as a result of interventions, specifically, intervention variables taking a specific value that fixes the value of the variable being intervened on (the difference between interventions and intervention variables will be discussed in subsequent sections). – There is absolutely more to say about variables, causal relata and representational devices that are used

to model the interaction between variables, but tackling these topics will be postponed until we have elaborated on the main theme of this section, the structure and conditions for causation (understood along interventionist lines, a topic we will return to and focus on now.

This concludes our remark about the first crossroad, reconciling the relativity of representation with the objectivity of dependency relations. We are bound to encounter more crossroads that overlaps with the preceding crossroad, but this is illustrating rather than fragmenting, as that allows us to highlight the central assumptions and ideas that influences interventionism as a whole.
We now turn to the central formulation of the interventionist conception of causation, the sufficient and necessary condition for X to be a contributing cause of Y, also known as, "the manipulability theory" (of causation).

6.3 The manipulability theory - Formulating a sufficient and necessary condition for contributing causation

what we are about to discuss is the heart of the semantics of interventionism, the manipulability theory, a sufficient and necessary condition for X to be a contributing cause of Y.

The manipulability theory consists of two sets of necessary and sufficient conditions, one describing what it is for X to be a (type-level) level direct cause of Y, the other describing what it is for X to be a (type-level) contributing cause of Y, both with respect to a variable set V. The set of conditions contributing causation incorporates the set of conditions for direct causation, such that the former is defined in terms of the latter.

In section 1.3 we gave a definition of direct cause (DC) that expresses the same idea of direct causation that occurs in the manipulability theory. However, they differ slightly in the sense that the latter introduces the notion of (type-level) causation (that stands in contrast to token-level), and that it specifies a variable Z_i when talking about all the other variables that needs to be held fixed (at some value). (Woodward 2003, page 59). – Therefore, we cite an independent formulation of direct causation that applies to the Manipulability Theory.

Definition of Direct cause (Manipulability Theory version).
"A necessary and sufficient condition for X to be a (type-level) direct cause of Y with respect to variable set V is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in V." – Woodward 2003, page 59.

As for characterizing the idea of contributing cause along the lines of the Manipulability Theory, we model the theory with the help of a causal model (that represents a causal structure with the help of a set of variables), M, consisting of the following variables; X (cause), Y (effect), Z_i ... Z_n (intermediate cause(s)), and P_i ... P_n (path(s)).

Definition of contributing cause (Manipulability Theory version).
For the necessary and sufficient condition for X to be a (type-level) contributing cause of Y with respect to variable set, V, the following criteria must be satisfied:

1. There is a directed path, P_i from X to Y such that each link in this path is a direct causal relationship. (That is, X must be a direct cause of every variable Z_i on the path P_i, of the set Z_i ... Z_n, that is one the path between X and Y, such that X is a direct cause of Z_i, which in turn is a direct cause of Z_1, which in turn is a direct cause of Z_n, which in turn is a direct cause of Y). (Woodward 2003, page 59).
   In other words, X must be a direct cause of every variable that links X and Y, such that there is a chain of direct causal relationships running from X, through Z_i ... Z_n, to Y.
2. There must be some intervention on X that will change Y when all other variables, V, that are not on this path (P_i), are fixed at some value. (That is, if there is another set of variables W_i ... W_n, on another path P_1(where $P_i \neq P_1$), that either links X and Y, or that leads into Y (by influencing Y through another path that does not go through X), there must be some intervention on X that will change Y, when all these off-path variables are fixed at some value). (Woodward 2003, page 59)

In the above formulation of the conditions for X to be a contributing cause of Y, we assumed that there is more than one path connecting X and Y, and that there are other paths that lead into Y (independently of X), but this is not necessary in order for the above conditions to be satisfied. X may qualify as a contributing cause "if there is only one path from X to Y or if the only alternative path from X to Y besides P contains no intermediate variables, provided that

there is some intervention on X that will change the value of Y, for some values of the other variables V" (Woodward 2003, page 59).

As we will mainly be concerned with the manipulability theory when accounting for the semantics of interventionism, it is worth to take a slight pause to elaborate on the truth-conditions of the manipulability theory, as these will play an important role when investigating the metaphysics of interventionism.

Woodward proposes two strategies for determining whether the truth-conditions for the manipulability theory is satisfied:

1. "Draw a causal graph that represents all of the direct causal relationships between X, Y, and all other variables in V. Check all of the routes or directed paths from X to Y. For every route P_i, freeze at least one intermediate variable at each of its various possible values along other routes from X to Y containing intermediate variables. Furthermore, fix all other direct causes of Y that are not one any directed path from X to Y at each of their possible value.
   If for some combination of values of these off-path variables, some intervention on X will change the value of Y, then X is a contributing cause of Y" (hence the truth-conditions of the manipulability theory are satisfied). (Woodward 2003, page 59-60)
2. Check whether there is some possible combination of values of all of the direct causes of Y that are not on the path of P_i (the directed path connecting X and Y), such that with these values fixed, some intervention on X will change the value of Y. In order for this strategy to work out, it is important that there is at least one path connecting X and Y, and that if variables that are not on this path are fixed at some value, changing X changes Y. If this operation succeeds, then X qualifies as a contributing cause of Y (hence the truth-conditions for the manipulability theory is satisfied). (Woodward 2003, page 60).

The common thread running through these strategies is that the evaluation of the causal relationship between X and Y proceeds by investigating the following aspects. (i) the direct route from X to Y, (ii) the alternative routes from X to Y (or other routes leading into Y), (iii) the effect of fixing the value of intermediate variables on other/alternative routes (if there are any). If changing X will change Y when other intermediate variables are held fixed, then X will qualify as a contributing cause of Y.

The fixation of variables is not arbitrary, as there is a measure for what (if any) variables should be held fixed. The right measure and fixation of variables is defined with respect to directed paths that runs into Y. As Woodward points out quite precise "(…) for each directed path between X and Y we are to hold fixed at each possible value at least on variable that is not on that path (or equivalently, all direct causes of Y that are not on that path) and then determine whether there is change in X that will change Y" (Woodward 2003, page 60)

As a way of summarizing the points from the preceding examination of the conditions and formulations of causation, while turning the focus to the topic that will concern us in the upcoming sections, we may say the following about the role and function of contributing cause (and its correlate total cause). Both these formulations of cause explicate and describe conditions having to do with what would happen under hypothetical interventions that are both necessary and sufficient for X to be a [insert designated form here] cause of Y.
To specify it further, what these two formulations aims at capturing, are facts about what would happen under hypothetical interventions (Woodward 2003, page 61).
Returning to a point that was made earlier about the importance of making sense of combinations of interventions, we may elaborate further on this point in the following sense. Integrating and developing the preceding idea further, the manipulability theory may be thought of as a codification of the idea that *a system of causal relationships* (a model consisting of variables that are causally relevant to each other) consists of a *distinctive set of causal claims* (about what would happen to the variables in the model under investigation), that expresses *facts* about what would happen under different combinations of hypothetical interventions (on the variables in the model), and given that there is an alteration in the *causal structure* (the pathway/routes/dependency relations among the variables in the model under investigation), then there will be an alteration in the content and truth-value of the claims about what will happen under interventions (hence, rendering the set of claims distinctive as each set depends on the causal structure). (Woodward 2003, page 61). Putting these observations together with the preceding discussions of the semantics of interventionism yields the following commitment for a manipulability account of causation: "No causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference" (Woodward 2003, page 61).

6.4 Causal claims, interventions, and domain restriction – restricting the assessment of causation to some, rather than all, interventions.

In the preceding section, we concluded with drawing attention to how the relationship between causation and interventions is cyclic in the sense that, the former figures in the characterization of the latter, and vice versa. In this section, we want to elaborate on this relationship, with particular focus on what commitments the interventionist analysis of causation implies imposes on the characterization of the scope, range and existence of interventions. These commitments are especially notable when it comes to elaborating on the difference between type- and token-level claims, while at the same time accounting for the generality of causal claims understood along interventionist lines.

First, Woodward notes that every condition that connects causation to interventions invokes the existence of "some" possible intervention on X that, under suitable conditions, changes Y (Woodward 2003, page 65). What we are interested in assessing is what these "suitable conditions" consists of, and what the connection between causation and the existence of "some" possible intervention amounts to for the elaboration of the truth conditions of the semantics of interventionism. – We name this point, "*the conditions of interventionist causation*" for future references.

Second, Woodward restricts the requirement that there exist possible interventions, such that it is necessary and sufficient that there is some intervention on X that would, under suitable conditions, change Y, rather than requiring that all interventions on X would change Y Woodward 2003, page 65). We are interested in assessing what this difference amounts to (whether it is a merely pragmatic choice, or if this carries substantial commitments), and how this is relates to the distinction between type- and token-level causal claims (that each invokes "some" possible interventions, albeit in a different manner and for different purposes). – We name this point, "*the domain of quantification for interventions*", for future references.

Third, Woodward draws a connecting between causal claims, in particular type-causal claims, counterfactuals and reproducibility. The idea is behind this connection comes down to the requirement that for a causal claim to come out as true, it must carry counterfactual import that supports the claim that if we were to repeatedly intervene on the causal relationship that figures in the causal claim, it would remain true under some of these interventions

(Woodward 2003, page 70).
What we are interested in assessing is what the interventionist analysis of counterfactuals implies for the interpretation of the truth conditions of counterfactual conditionals in relation with type- and token-causal claims (in particular, the truth of type-causal claims, and the evidential support of token-causal claims). (Woodward 2003 page 72-73).
Furthermore, we are interested in elaborating on the connection between the objectivity of counterfactuals conditionals, the relativity of counterfactual reasoning (in particular, the assessment of which possibilities we take seriously), and the implications this connection has on the metaphysics of interventionism. – We name this point, "*the metaphysics of interventionist counterfactuals"*, for future references.

As to the first point concerning the "suitable conditions" for causal claims and their related causal relationships to come out as true, we may use one of Woodward's own examples to illustrate what "suitable conditions" are meant to capture. According to Hooke's law, the restoring force F, exerted by a spring s is described by the following equation: $F = -k_s X$, where X is the extension of a particular type of spring and F the restoring force it exerts. Considering that the restoring force F is a function of the extension X, and that it is possible to change the value of F by intervening on X, Hooke's law describes a causal relationship between the extension and the restoring force. Hooke's law illustrates the idea of "suitable conditions" by carrying information about the conditions of where the causal relationship will hold (possible to intervene on X to change F), the conditions of how and to which extent the value of the variables in the causal relationships influence each other (How does the manipulation of the value of X influence the value of F), and, the conditions of where the causal relationship will break down (not possible to intervene on X to change F).

The idea that motivates the relativizing (or restriction to be more precise) of Hooke's law to "suitable conditions", is that the law will have exceptions due to the inability to intervene on X, or the presence of additional/other background conditions that prevents the causal relationship from holding (Woodward 2003, page 68). – The focus is not on Hooke's law per se, but rather the logic of interventionist semantics where suitable conditions are connected to the presence of factors in the causal model/claim under investigation and the "ability to intervene on the variables that figures in the causal model/claim" (what ability to intervene consists in will be discussed in detail later). Hence, what the logic of interventionist semantics prescribes for the analysis of causal claims, such as Hooke's law, is that we should be restrict

the assessment of causal claims to what would happen under *some range of manipulations of X, for some, but not all springs* (Woodward 2003, page 68).
To summarize the criteria that "suitable conditions" imposes on the interventionist semantics:

As to the second point concerning "the domain of quantification for interventions", once we have restricted the assessment of causal claims to "suitable conditions", it seems to imply that the domain of quantification for possible interventions on X is restricted in a similar manner. As suitable conditions and domain of quantification complements each other, there will be, and indeed is, some overlap in both their roles and definitions. But, as a way of demarcating (at least temporarily) between the two, we might say that "suitable conditions" plays the role of truth conditions that provides information about the content and truth-value of causal claims, whereas "domain of quantification for interventions" concerns the existence, influence, and metaphysics of interventions, and hence provides information about the range and possibility of utilizing interventions for causal assessment (of both causal claims and relationships).

As a way of setting up the discussion of the domain of quantification for interventions, we may draw attention to a related distinction that Woodward draws between two types of information about causal claims; 1. Information about there being some intervention on X that will change Y, hence render the claim X causes Y true, and 2. Information about which interventions on X that will change Y, under what circumstances those changes will occur, and finally, how these interventions on X will change Y (Woodward 2003, page 66). For future references, let's call 1. Generic causal claims, and 2. Specific causal claims.

To illustrate with an example; the distinction comes down to the difference between the claims: there is some intervention on the smoking habits of some group of individuals that is part of a population that would change whether the group would develop cancer, and hence render the claim "smoking causes cancer" true. In contrast, if we would have tested two groups of people, one control group and one experimental group, intervened on the smoking habits of the experimental group by increasing their intake of cigarettes during a lengthened period of time, while holding fixed other variables that might influence the results, such as natural exposition to smoke from traffic, then we would see a corresponding increase in the development of cancer in the experimental group (as opposed to the control group).

The difference between the two claims might seem familiar, and there is a reason for that; the difference corresponds to the difference between type- and token-causal claims that was introduced in the definition of the manipulability theory M (See section 1.4).
What the difference between the two types of claims consists of, is a question that addresses the intricate relationship between causal information expressed by type-causal claims of the form, X is causally relevant to Y, token-level causal claims of the form Y is a function of X (under these circumstances and in this specific sense), and the requirements that the two types of claims impose on the truth conditions for causal claims. – We started with posing the question of how generic- and specific causal claims relates to type- and token causal claims, and illustrated that the relation is a matter of the degree of causal information that the two pairs and two types of claims expresses. Now that we have proved a connection between the two pairs of causal claims (generic-specific, type-token), the next pressing question is what the difference between the two types of causal claims consists of (generic-type versus specific-token), and whether this difference is a matter of different level of analysis and/or if it is such that one of the claims are somewhat more "basic" than the other (and if so, how and to what extent this influences the assessment of the metaphysics and truth conditions of causal claims)?

Before turning to discussing the relation between interventions, truth-conditions and type-contra token-causal claims, I want to turn the attention to a second distinction that illuminates the underlying logic of interventionism; the distinction between internal and external validity of experiments (and their corresponding causal claims).
This distinction is made in the literature on experimental design, and the idea is that by drawing the distinction, we might characterize the conditions under which some experiment and causal claim will hold. This is done by specifying the range of contexts and background conditions under which the causal relationship holds, and its corresponding causal claim comes out as true, while supporting the assessment with the help of experiments (where truth is tied up with the assessment of the truth conditions of the experiment).

According to Woodward; "the internal validity of an experiment has to do with whether the claim that *some factor C causes some effect E* in *the background conditions* characterizing *that very experiment* is true." (Woodward 2003, page 69: my own emphasis). – The underlying idea is that there should be some causal relationship, that holds in some background condition, such that the experiment reveals the causal connection under the

specified conditions. In other words, what it means for an experiment to be internally valid, is for there to be a specific context and causal relationship in which the experiment provides insight into the truth of the claim that C causes E.
Furthermore, and by contrast, "external validity has to do, roughly, with the extent to which the causal relationship between C and E will *continue to hold in other circumstances* besides those obtaining in the original experiment – with the extent to which the C-E causal relationship *generalizes to other subjects and background circumstances*." (ibid: my own emphasis). – Here, the idea is that for some causal relationship, there should be a set of background conditions, besides the one where the experiment is performed, such that it is possible to extrapolate and generalize the causal relationship in a way that it holds in those background conditions (including the one where the experiment is performed). In other words, what it means for an experiment to be externally valid, is for there to be a set of contexts where the causal relationship holds that goes beyond the one context where the experiment is performed.

As a way of drawing the connection between the two forms of validity, we might characterize validity as a feature having to do with the domain of experimentation, where the difference between the two comes down to the range and conditions under which the causal relationship holds (the one is restricted to the conditions where experimentation is performed, while the other holds outside of these conditions (including the one where experimentation occurs)). - There is a striking familiarity between domain of experimentation, and the notion we introduced earlier, domain of interventions, which is natural considering that interventions serves the role of an experiment (if we by experiment mean an operation used for causal assessment), while experiments (at least in this context) serve the role of causal assessment, just as interventions does. Hence, we will subsume talk of domain do experimentation under domain of interventions.

Putting the remarks explicating the distinction between generic-specific causal claims (and their relation to type-token causal claims) together with the remarks relating to the distinction between internal and external validity, while coining this together with our discussion of interventions, truth conditions, suitable conditions, and domain of quantification yields the following predicament.

Within the interventionist framework, each of these notions and topics are explicated with the help of another central notion, namely that of invariance under intervention, the idea that for

every causal relationship/claim that holds (that it is true that there is a causal relationship C-E, some intervention on C, and some background circumstances, where E change under interventions on C) there will be some background circumstances where the relationship/claim under question will continue to hold/not break down if intervened on (or if other changes occur as will be discussed later). (Woodward 2003, page 69).
In other words, what invariance under intervention implies, is that there is some feature of the causal relationship/claim that would remain unchanged (the fact that the causal relationship between C and E holds), whenever changes in the background conditions and/or the causal relationship/claim would occur (such as intervening on C, or holding other variables Z, fixed at some value). Whenever a causal relationship possesses this feature, it is said to be invariant under the interventions and background circumstances under investigation. – The central role of invariance within the interventionist framework reveals itself in the way that the interventionist conception of causation is tied up with the notion of invariance, such as the following statement expresses; "(…) if a relationship is to qualify as causal, it must be invariant under some interventions" (Woodward 2003, page 69).

Furthermore, it should be noted that invariance exhibits and implies the same sense of domain restriction and relativity as interventions, causal relationships and counterfactual conditionals discussed earlier. When assessing the invariance conditions of causal relationships, we restrict the assessment to some, rather than all, interventions, while at the same time relativizing the assessment to a set of interventions and background circumstances.
What this implies, is that a relationship may be invariant under some interventions and background circumstances but not invariant under others, and in turn, that the content of causal claims might be spelled out by describing the range or domain of interventions and background circumstances over which it is invariant (Woodward 2003, page 70).
The current characterization of invariance is only a preliminary account of how invariance connects to the discussion of causation within the interventionist framework. The notion will receive more extensive treatment in subsequent chapters, where the focus will be on the logic, metaphysics and implications of invariance under intervention.

To sum up the central points concerning the relationship between causation and invariance: For a causal claim to be true, there must be some hypothetical intervention on the causal relationship, C causes E, that figures in the causal claim that satisfies the following criteria:

1. If the intervention were to be carried out, such that intervening on C would change E (stating a truth-condition for the claim).
2. The causal relationship would have to continue holding under that intervention, in other words, remain invariant under that intervention (hence, satisfying the truth-condition for the claim)
3. Provided that the above criteria are satisfied, the causal claim "C causes E" would be true, as there would exist an intervention on the causal relationship figuring in the causal claim that would remain invariant under intervention. – Hence, the invariance under intervention figures as the central constituent in the truth-conditions for causal claims.

Now that we have established the connection between causation and invariance, while highlighting how the restriction of causal assessment to some, rather than all, interventions establishes a close connection between truth-conditions and invariance conditions, we are in a position to elaborate on one of the other central aspects of interventionist semantics, the connection between interventionist causation and counterfactuals.

6.5 Causation, counterfactuals and reproducibility – generalizing invariance conditions.

An understated aspect of the interventionist semantics, is the connection between causal claims and counterfactuals. Both formulations of causation (TC) and (M) invoke counterfactual conditionals when elaborating on the truth conditions of causal claims. According to (NC*), one of the conjuncts that makes up (M), if X is a contributing cause of Y, then the following counterfactual conditional will be true: there is some intervention on X, such that, if other variables were held fixed, Y would change. Furthermore, (SC), the other conjunct making up (M), embodies the following counterfactual: if an existential claim about what would happen under a counterfactual supposition is true (that is, if there is a possible intervention on X that would change Y), then X causes Y. Woodward 2003, page 70). This implies the following predicament for (M) and (TC), both of them qualifies as counterfactual theories of causation as they integrate a systematic connection between causal claims and counterfactual conditionals (ibid).

The theme of this section, reproducibility, is closely connected to the interpretation of counterfactual conditionals of the previous sort(s). For the assertion "there is some possible intervention on X that would change Y" to be true, where X and Y are causally relevant to each other (there being some change in the value X that would influence the value of Y), and where causal relevance is a type-causal notion (there being some range of values for the changes in X and Y, rather than a particular value in a particular context for the changes in the two variable), is for there to be a reproducible relationship between X and Y. (Woodward 2003, page 40, 41). – More will be said about how exactly interpretation of counterfactuals hinge on the notion of reproducibility.

Furthermore, Woodward makes the explicit restriction to accounting for causation in deterministic contexts (with some side remarks about indeterministic causation), which in many ways reinforces and expands the role of reproducibility in the interventionist account. Understood along interventionist lines, what this restriction implies, is that "if X is a deterministic cause of Y, then a set of associated counterfactuals specifying how Y would change under manipulations of X (and possibly other variables as well) will be true)." (Woodward 2003, page 41). As a way of lifting this restriction, so as to encompass an account of indeterministic causation, it is possible to reformulate the "set of associated counterfactuals", such that they are formulated in terms of probability distributions (rather than changes in values) (ibid).
As was described earlier, the basic relation of causal relevance between variables, implies the existence of reproducible relationships, which is another way of stating that there is a systematic relationship between the changes in the value of X, the response in the value of Y, and the repeated interventions on X (under the right circumstances/conditions).

Reproducibility combined with the restriction to determinism, has the following implication for the causal relevance between X and Y: "In the deterministic case, Y *always changes* in the *same way* under the *same kind* of manipulation of X, at least for *some range of manipulations* for individuals within *some range of background circumstance*s" (Woodward 2003, page 42, my own emphasis).
This conjunction is an implication of erecting the interventionist analysis on the top of the notion of type-causation (as Woodward himself states; ibid)), which in many ways centers the idea of reproducibility in the heart of the interventionist framework.
As a way of adjusting the conjunction such that it captures the idea of indeterministic

causation, it is possible to replace the one conjunct (determinism with indeterminism), while replacing talk of sameness of manner and kind of change, with sameness of production and influencing on the probability (distribution), which in turn yields the following account of the causal relevance of X and Y: "When causation is indeterministic, reproducibility instead may be understood to cover a range of possibilities: it may be that for some range of background circumstances, the same kind of manipulation of X always changes the probability of Y by the same value or, more weakly that repetition of a kind of manipulation of X in the same background circumstances always produces either an increase (or decrease) in the probability of Y, although perhaps by different amounts on different occasions." (Woodward 2003, page 42). – The focus will remain on the deterministic contexts and formulations of the interventionist framework, while the indeterministic contexts and formulations will remain as an illustration of the flexibility and scope of reproducibility (and invariance).

Hence, what the discussion of reproducibility amounts to at this point, is that the role and characteristics of the counterfactual commitments of the interventionist analysis, the dichotomy between type-token causal claims, and the difference between deterministic and indeterministic contexts all point in the direction of reproducibility serving as a central feature of interventionist semantics.

As pointed out earlier, there is more to say about the details of the relationship between reproducibility and counterfactual conditionals than the mere fact that it plays a central role in its elucidation, a topic that I know turn to. As a preliminary note, I propose that the role of reproducibility in the evaluation and interpretation of counterfactuals (hereby assessment of counterfactuals for brevity) can be divided into three parts: 1. What it implies for the assessment of counterfactuals (that there is a reproducible type-causal claim underlying it), 2. What it rules out for the assessment of counterfactual claims (type-causal claims with singularist content and singularity as part of the meaning of causal claims), 3. What it amounts to in the assessment of counterfactuals (truth-conditions and evidence for counterfactual claims).

1. What reproducibility implies for the assessment of counterfactuals

The implication of reproducibility for the assessment of counterfactuals, particularly, the counterfactual claim that, there is some possible intervention on X that would change Y, understood along the lines of both (M) and (TC), is that the claim should be understood as

expressing the following: "there is some intervention on X such that if it were possible to intervene to manipulate X repeatedly in that way, Y (or the probability of Y) would change in some reproducible or repeatable way" (Woodward 2003, page 70-71). – This implies that the interventionist analyses (M) and (TC) being implicit causal generalizations, where the meaning of 'generalization' is that there is a systematically general/reproducible relationship between X and Y, under which, if it were possible to manipulate X repeatedly, Y would change in some repeatable way, and where the generalization would hold for a range of circumstances under a number of occasions.
As a way of further elaborating on the commitments of reproducibility, we might think of it as "the idea that the response of Y (or the probability of Y) to manipulation, should be general or stable enough that it is appropriate to speak of manipulating X as a means or strategy for producing Y in some (perhaps highly restricted) circumstances" (Woodward 2003, page 71)[10]. Hence, what might be concluded when it comes to the commitments and implications of reproducibility on the assessment of counterfactual claims, is that these claim are both implicitly and explicitly restricted to reproducible causal generalizations.

As a note, it is important to remark that despite discussing and elaborating on reproducibility with the help of ideas such as, range of circumstances, and systematic connections between variables (more specifically, the response of variables to interventions), it is not built into the idea of reproducibility that causal generalizations have external validity or that the relationship between the variables is deterministic (Woodward 2003, page 71).

As we might recall, the difference between internal and external validity had to do with the circumstances under which the causal relationship would hold, and where external validity designated the attribute that a causal relationship holds under other circumstances (including the one where the experiment is performed). Reproducibility, although connected to this idea, does not commit us to suppose that a causal relationship has to be externally valid (ibid).
A causal relationship that is reproducible in a restricted and specific experimental context, will qualify as a reproducible regardless of whether it holds under different contexts or

[10] The idea of linking reproducibility and causation to means/strategies hinges on the idea Cartwrights idea of thinking of causation as "effective strategies for realizing certain goals", which does indeed have both meta- and substantial philosophical implications for the epistemology and metaphysics of causation, which will be discussed in detail when treating the metaphysical picture that underlies Woodward's formulation of interventionism.
To hint at the conclusion, the idea of thinking of causation in methodological terms (as means for certain ends), generalizes to the conception of metaphysics that underlies interventionism, namely that, the metaphysical commitments and implications is sensitive to the means and ends of our inquiry.

circumstances.
The same goes for the difference between deterministic and indeterministic relations among variables. If the former denotes a relationship where interventions always change the value of variables in the same way, and in the same manner, while the latter denotes a relationship where interventions always produce the same change in the probability distribution of the variables, then both relationship will qualify as a reproducible relationship. In other words, reproducibility does not commit us to suppose that causal relationships have to be deterministic (ibid).

2. What reproducibility rules out for the assessment of counterfactual claims

In the preceding section, we highlighted how reproducibility, by being an integral part of both (M) and (TC), introduces a number of commitments, the central one of them being that there are reproducible causal generalizations underlying the assessment of counterfactuals.
What we want to in this section is examining some of the implications that follows from these commitments, particularly, what these commitments rules out as viable candidates for assessing counterfactual claims.
To hint at the conclusion, in light of the fact that causal generalizations underlies the assessment of counterfactual claims, causal singularities, causal claims that has implications only for what happens in a single instance as part of their content and meaning, are ruled out as sufficient conditions for the truth of counterfactual claims (Woodward 2003, page 71-72). – In other words, reproducibility introduces significant restriction on the truth-conditions for causal claims, a topic which I now turn to elaboration upon.

As was described in the previous paragraph, there is a distinction between generality and singularity, the two terms corresponding to the difference between type- and token-level causal claims, the former being central to the assessment of counterfactual claims, the latter being ruled out as a legitimate candidate for assessing counterfactuals claims.
There is a further distinction to draw between two senses of singularity, singularity as constituent of the meaning of causal claims (semantic singularity), and singularity as a contingent fact (metaphysical singularity).

Semantic singularity relates to the truth conditions of counterfactual claims in the following sense; when semantic singularity is *part of the meaning of counterfactual claims*, the following counterfactual statement, "there is some intervention on X that would change the

value of (or the probability distribution of) Y", *would be satisfied by the following condition* "there is some single token intervention that changes the value of X and under which value of Y (or the probability of Y) also changes", *such that the counterfactual claim would come out as true* (Woodward 2003, page 71-72).
In other words, semantic singularity is the thesis that the truth-conditions of counterfactual claims are satisfied by, and that the meaning of counterfactual claims are built around: token interventions (interventions that are single instance(s), occurs at a single context(s), and for a particular causal relationship). This sense of singularity is ruled out by the condition of reproducibility as it restricts the meaning of counterfactual claims to singular .

Metaphysical singularity relates to the truth conditions counterfactual claims in the following sense; when singularity is a contingent fact, there might be a counterfactual statement, "there is some intervention on X that would change the value of (or the probability distribution of) Y", that is satisfied by a condition "there is some single token intervention that changes the value of X and under which value of Y (or the probability of Y) also changes", and where the counterfactual claim comes out as true, not by being built into the meaning of the counterfactual claim, but rather as a contingent fact about the world (where the intervention happens to be realized at a single instance, at a single context, and for a particular causal relationship in the history of the world).

In other words, metaphysical singularity is a thesis that describes the conditions under which the counterfactual claim comes out as true (by describing the state of the world, and, the moment when the variables take the designated values that realizes the causal relationship), but it does not specify the content (of the counterfactual claim) and systematic relationship among the variables figuring in the counterfactual claim (as that it part of the meaning of counterfactual claims). (Woodward 2003, page 73)
Hence, metaphysical singularity is not ruled out by reproducibility, as it is possible (in principle), that the causal relationship between X and Y could be realized at another occasion, and the circumstances under which the counterfactual claim comes out as true could be replicated, which does not contradict the requirements laid down by reproducibility (that a counterfactual claim would have to be reproducible in order to come out as true).

The difference between the two sense of singularity, semantic and metaphysical, seems to reveal that Woodward operates with (at least) a two-fold notion of truth conditions for counterfactual claims, the first fold specifying *how* the counterfactual claim would come out

as true, the second fold describing *when* (*the circumstances under which)* the counterfactual claim would come out as true. –Whether, and if so, how, this distinction influences the metaphysical commitments of interventionism are questions that will have to be postponed to answer until the section that discusses the metaphysics of interventionism in general[11].

3. **What reproducibility amounts to for assessment of counterfactuals (truth-conditions and evidence for counterfactual claims**)

At this point, we have highlighted the commitments and exclusions that reproducibility imposes on the assessment of counterfactual claims (that there are causal generalizations that underlies them and that causal singularity is ruled out as part of the meaning of counterfactual claims). In this section we want to discuss what this amounts to in the assessment of counterfactual claims, particularly, how it influences the truth-conditions of, and evidence for counterfactual claims in general (including both type- and token- causal claims).

In the preceding section, we presented an extensive number of arguments and considerations that cements the position of reproducibility and causal generalizations at the foundation of the truth-conditions for counterfactual claims, so rehearsing these arguments here would be redundant. However, what haven't been discussed for so far, is the extent of which the reproducibility and causal generalizations applies as truth-conditions, in particular, what other forms of causal-claims have them as indispensable parts of their truth conditions.

So far, it should be clear that type-causal claims integrate reproducibility as a part of their truth-conditions, but what may not be so clear, is what this means for token-causal claims. Woodward contends, and I agree with him, that "token or singular causal claims always should be understood as committing us to the truth of some type-level causal generalizations", which is a consequence of the fact that "(…) we must appeal to claims about type-causal relationships of the sort embodied in (M) to elucidate token-causal claims (…)" Woodward 2003, page 72).
As my focus is restricted to type-causal claims, I will not argue in detail about the interventionist conception of token-causal claims, but what can be said to substantiate the preceding claim, is the following.
According to Woodward, "(…) information about deterministic type-causal relationships is

[11] Consider whether to discuss this in light of Tooley "Causes, laws and ontology" (The oxford handbook of causation) to reveal the intricate details of the singularity versus generality discussion, and its implications for the ontological commitments of the debate.

assumed to be part of our background knowledge, and the only question is what these type-causal relationships and other background information imply about token-causal relationships (Woodward 2003, page 75). Combining this assumption with the idea that token-causal claims involves the combination of interventions (fixing variables at specific values) and variables taking actual values (under intervention), we may conclude that type-causal relationships play a fundamental role in elucidating actual-causal relationships, which is a consequence of both of them embodying a counterfactual interventionist conception of causation (Woodward 2003, page 79).
As a concluding note, in light of the preceding remarks and considerations, reproducible type-causal claims function as truth-conditions for type- and token-causal claims.

As to the second implication of requiring reproducibility (for the assessment of counterfactual claims), the relation between truth conditions, and realization of conditions, specifically, how the two relates to the evidential basis for assessing counterfactual claims leads to the following predicament.
As a way of balancing between; on the one hand, the generality of truth conditions and meaning of counterfactual claims, and, on the other hand, the singularity of the antecedents realizing the truth conditions of counterfactual claims (conditions consist of interactions between variables/causal relationships instantiating spatiotemporally located individuals/events[12]), the evidential basis for assessing counterfactual claims seems to require a partial consolidation of the type-token distinction if it is to succeed in providing a test procedure for assessing the truth of counterfactual claims. – What the evidential basis consists of, how it relates to the predicament just posed, and whether it is possible to overcome it while respecting the type-token dichotomy, are the topics of the rest of this section.

[12] This description of the content of conditions is only partial, as a complete specification of the content would include more details, such as the dependency relations, range, and scope (of the dependency) among the variables under question.
The current description of conditions refers to Woodward's own conception of what the claim "X is causally relevant to Y" (the standard formulation of type-causal claims) consist of, particularly, what it's truth conditions consist of: "In my view, a claim such as "X is causally relevant to Y" *is a claim to the effect that changing the value of X instantiated in particular, spatiotemporally located individuals will change the value of Y located in particular individuals. Thus, the truth of a claim such as (S) "Smoking causes lung cancer" depends on relationships that do or would obtain (under appropriate manipulations) at the level of particular individuals*, even though it is also true that (S), as it stands, does not tell us anything about the causes of any particular episode of cancer (…) or indeed that any individual either smokes or develop lung cancer" (Woodward 2003, page 40). – I extend this view to analyze the different components of the truth conditions of counterfactual claims, and particularly, the evidential support and basis of the truth of such claims.

The question of what the evidential basis of counterfactual claims consists of can be answered in two parts; 1. The state of the causal correlation (figuring in the counterfactual claim) once the condition specified in antecedent (of the claim) is *realized*, 2. The result of *carrying out* the interventions described in the antecedents of the claims (Woodward 2003, page 72). – The two notions written in italics, 'realized' and 'carrying out', are the sources of contention in the evidential basis, as they, contrary to the hypothetical interventions and causal generalizations (figuring in the counterfactual claims and their truth conditions), have a factual (rather than hypothetical) and specific (rather than general) character.
It is these two characteristics, those of factuality and specificity, that consolidates between the type-token distinction, as key characteristics of the correlations figuring in the evidential basis reminisces of token-causal claims, rather than type-causal claims, due to their factuality (taking actual values under interventions that are carried out) and specificity (restriction to specific conditions rather than a range of circumstances), whereas the evidential basis is used for assessing the truth of counterfactual, type-causal claims.

Expounding on the seeming consolidation, we can say that the reproducible causal generalizations figuring as truth conditions, make up the semantic content of counterfactual claims (remember that semantic singularity is ruled out, which leaves causal generalizations as the only candidate), while the causal (token) correlations figuring in the evidential basis, make up the empirical content of counterfactual claims (remember that metaphysical singularity is still a live option as it is contingent on how the world actually is, which serves as an illustration that there is no contradiction between simultaneously accepting singularity as a contingent matter and generality as a semantic matter).

As to the final question of how the (token) correlations in the evidential basis for, and the (type) generalizations in the meaning of counterfactual claims relate to the type-token dichotomy, there is an ingenious answer to the above predicament. – The (token) correlations figuring in the evidential basis function as evidence for the truth of counterfactual claims, but they are neither entailed by, nor entail the truth of, the causal generalizations figuring in the counterfactual claims (Woodward 2003, page 73). This is a considerable concession, as the distinction between entailment and evidence allows Woodward to uphold the type-token distinction, the primacy of type-causal claims, while cementing the empirical component of the assessment of counterfactual claims.

As a way of tying the loose ends and summarizing the main points concerning the evidential basis of counterfactual claims, we might say that the presuppositions and truth-conditions of (M) and (TC) invokes existential claims about possible (hypothetical) interventions, whereas the evidential basis of these counterfactual claims refer to actual interventions in the elucidation of what it means for the truth-conditions to be satisfied (in other words, what the evidence for the counterfactual claims are). – As to the question of what kind of implications this has in terms of metaphysical commitments, and furthermore, how this relates to the metaphysics of interventionism in general, are questions that will be discussed in due time.

Before turning to discussing the content of counterfactual conditionals, particularly, the counterfactual possibilities that form the range and background for assessing the truth of counterfactual claims, it should be noted that the above distinction between the truth conditions of, and evidence for counterfactual claims is subject to idealization.
As Woodward notes, there are two ways in which idealization figures in the above discussion: 1. The assumption that, "the units in the correlations that was experimentally manipulated share the same causal structure", 2. The assumption that, "the values of X and Y are observed and/or measured for all units that are experimentally manipulated" (Woodward 2003, page 74). By using idealization, Woodward is able to ignore the issues having to do with selection bias and missing measurements, issues that have to do with practical considerations.
The presence of idealization plays a considerable role in the assessment of counterfactual claims, as that introduces considerations having to do with, not how the world is, as in the strategy of assessing the (token) correlations in the evidential basis, but rather, considerations having to do with how we represent and acquire knowledge of the world. The latter considerations are methodological (and partly epistemological), which renders it a thorny issue whether we can read off metaphysical commitments from the interventionist analysis of causation, as the analysis itself is subject to considerations not having to do with metaphysics, but rather methodology. – This issue, and its implications will be discussed in detail under the heading of metaphysics versus pragmatics in the interventionist framework.

6.6 Causation, omissions and serious possibilities – restricting the range of counterfactual possibilities and background conditions

Earlier, when discussing the notions of direct causation, variable sets, and causal pathways, we came across the contentious topic of variable choice, the idea that our representation, and hence, assessment of causal claims is subject to how, and what variables we include in our causal models (Woodward 2003, page 53, 56-57). This idea seemed to threaten the objectivity of causation by introducing considerations having to with our interests and beliefs about variables, rather than the dependency among the variables.
We presented a solution to the contention, and a way of salvaging the objectivity of causation by restricting the relativity/subjectivity to the representation of causal relationships (hence conceding that pragmatic considerations do play a role in a theory of causation and explanation), but upholding that when the representation of the causal relationships has been fixed, and the variables chosen, it is an objective matter whether the causal relationships between them hold (or breaks down).

However, we did hint at a further contention in the above discussion, a contention the threatened to deepen the relativity of the interventionist analysis, a contention that we postponed discussing, until now. The contention I'm referring to is the other side of the coin of variable choices. One thing is the idea of variable choice and the pragmatic considerations it introduces, another thing is what the choice of variables represent, which is, "(…) those possibilities that we are willing to "take seriously"" (Woodward 2003, page 56-57).

The idea of serious possibilities, threatens to undermine the objectivity of causation (despite the above solution), as now the contention is not on the side of representation (as in variable choices), but on the side of the counterfactual dependencies (as the truth-value of the dependency relations will be affected by which counterfactual possibilities are viable).
How Woodward solves this contention, how it is possible to reconcile the objectivity-subjectivity dichotomy, while recognizing the existence of serious possibilities, and what kind of implications this has for the metaphysics of interventionism will be the topic of this section.

To hint at the conclusion, just as the above discussion about variable choices led to the need of conceding to the pressure, thereby demarcating between the domain where relativity/subjectivity necessarily applies, and the domain where objectivity is safeguarded, a similar concession is needed in the discussion about serious possibilities.
The question is not if concessions have to happen, rather, the question is what kind of

concession and to what extent does an interventionist analysis of causation have to concede? Let me tackle these questions and topics in turn.

Woodward remarks the following concerning the basis for evaluating whether a conditional represents a serious possibility: "Obviously, the extent to which an outcome represents a serious possibility is a matter of degree, to which a number of considerations are relevant" (Woodward 2003, page 88).
Therefore, when discussing the idea of serious possibility, what I will do when trying to highlight the kind and extent of concessions that the notion introduces, is to identify what these 'considerations' are, and whether they are subjective or objective (where the distinguishing factor between the two is whether the consideration is seen relative to our interests/choices or the world).

At the outset of the discussion, the two worries that an account of causation that admits serious possibilities have to tackle can be divided into two: the threat of subjectivity (that the choice of serious possibilities is an interest-dependent choice) and the threat of unclarity (that it is unclear how, under what circumstances, and with which criteria we judge a possibility to be serious).

Woodward responds to the threat of subjectivity in a similar manner as, when he dealt with the threat of relativity when discussing variable choices; by distinguishing between the cases and circumstances in which subjectivity applies, from those where it fails to apply, while demarcating between the considerations that are sensitive to how the world is, from those who aren't (Woodward 2003, page 89). – The subtle transition from talking about cases and circumstances in which subjectivity applies/doesn't apply, to considerations that are/aren't sensitive to the world is not arbitrary, and it reflects an underlying distinction that serves as tool for demarcating the domains of subjectivity and objectivity.

The underlying distinction I have in mind is between causal judgment, which on the one hand, concerns how investigators evaluate and assess serious possibilities, and counterfactual structure/dependence, which on the other hand, concerns the relationship among variables that are causally relevant to each other (in other words, the dependency relations figuring in causal relationships). – As have been pointed out at earlier times, the interventionist analysis commits itself to the objectivity of causation, which in this case is reflected in the way counterfactual dependencies hold independently of which possibilities we take seriously, and

hence "(…) are interest-independent and objective" (Woodward 2003, page 90).
Causal judgment consists of a two-layered structure, where the subjectivity and interest-dependency serious possibilities, constitutes a layer of the basis for causal judgment, whereas the objective and interest-independent patterns of counterfactual dependence constitutes the other layer. – Hence, the subjectivity is restricted to the evaluation of which possibilities to consider as serious, but once these possibilities are fixed (once we relativize our assessment to a set of serious possibilities), there is no further element of subjectivity nor interest-dependency that influences our judgment about the truth-value of causal claims, as this will be determined by objective patterns of counterfactual dependence (Woodward 2003, page 90).

At first glance, this proposed 'two-layered structure' seems to introduce a considerable amount of subjectivity in the sense that we have to relativize our causal judgment to a set of serious possibilities in order to assess the truth-value of causal claims, and where these serious possibilities will be subject to our interests and evaluation. – But, there is more to the story than what the glance seems to imply.
As Woodward remarks, "although counterfactuals concerning nonserious possibilities can be straightforwardly true or false, to the extent that the possibilities that figure in them are nonserious, they do not guide our causal judgments. Instead, our causal judgments will be influenced just by those counterfactuals we think are true and concern only serious possibilities." (Woodward 2003, page 90) – The idea that is expressed in this passage, is that the truth-value of causal claims is independent of our causal judgments, as counterfactuals that concerns nonserious possibilities "can be straightforwardly true or false (…)" (ibid).
The fact that counterfactual claims can have a determinate truth-value regardless of whether they represent serious possibilities, makes them subject to objective, rather than, subjective considerations.
Putting this point together with the preceding remarks concerning counterfactual dependencies, translates to (at least) two objective constituents in the assessment of counterfactuals: the truth-value of counterfactual claims and the counterfactual dependency relations (Woodward 2003, page 90, 118).

There is a further implication of the above passage concerning the relation between causal judgment and possibilities, namely that, the manner in which causal judgment is restricted to those counterfactuals 'we think are true', seems not only to cement the role of subjective consideration in our causal judgment, but it also seems to introduce an epistemological

dimension to the assessment of counterfactuals. Causal judgment seems to have an epistemic dimension that concerns how we judge and infer about which counterfactual possibilities to consider as serious, whereas counterfactual dependencies/truth-values of counterfactuals seems to have ontological dimensions that concerns how the world/causal structure is. – This is not surprising as, for reasons to be discussed later, in Woodward's interventionist framework, the relation between epistemology and ontology is intertwined in the sense that the former constrains the latter (as a consequence of the emphasis on epistemic accessibility), whereas the latter has primacy over the former (as a consequence of the objective and realist underpinnings of the former).

The fact that causal judgment is constituted by two layers, and where the weighing of which counterfactual possibilities to regard as serious (with its subjectivity and interest-dependence) comes apart from counterfactual dependencies (with its objectivity and interest-independence), is an important maneuver that Woodward uses to salvage the objectivity of causation, while encompassing the subjectivity of causal representation (variable choice, serious possibilities).

As a way of turning the attention to the next chapter in the discussion of interventionism, while tying it together with the preceding discussion, I will argue that, the commitment to realism about causation (the thesis that there exist mind-independent/objective dependency relations and truth-values for counterfactual claims) introduces an explanatory and metaphysical asymmetry between the subjective and objective aspects of the interventionist framework, in which the subjective considerations stands in an asymmetric relation to objective considerations.
In this case, the asymmetry reveals itself in the manner of which causal judgments are evaluated on the basis of their correspondence with the counterfactual dependencies that hold in the system under investigation (and not vice versa).
What this 'realism about causation', and particularly the explanatory/metaphysical asymmetry consists of, and how this relates to the characterization of interventions, a notion that have been used loosely throughout this entire discussion and which figures in the foundation of interventionism is the topic of the upcoming section.

7 Interventions, variables and metaphysics – Getting to the bottom of the ‘nature’ of interventions

In the two preceding chapters, we provided a preliminary account of what interventions consists of, their role in causal query, and finally, the structural connection between interventions and causation. The picture we drew of interventions depicted them as hypothetical/ideal experiments that serves the role of causal assessment, namely, ideal experiments that determines whether and in what sense a correlation between two variables, X and Y, qualifies as causal.
Furthermore, we highlighted some of the suppositions underlying interventions, particularly, that in virtue of being hypothetical, it is not required that interventions are practically or physically possible, that the modal commitments of interventions are connected to variables and facts about counterfactual claims, and finally there are different ways of characterizing interventions (formal and informal).

What we are aiming at accomplishing in this chapter is not only expanding this picture, but also enhancing its quality by bringing out its apparent and hidden details. – Formulated less metaphorically, what we aim at accomplishing is highlighting the theoretical characteristics of interventions, with an emphasis on its formal definition(s) and suppositions, and furthermore, elucidate on the logic of interventionism, in particular, the relationship between the epistemic, ontic and pragmatic features of interventions.
As a way of accomplishing these two tasks, we are particularly interested in posing, and in turn, answering the three following questions; what exactly are interventions, and how does the nature of interventions influence the logic of interventionism? What is the connection and difference between interventions as causal assessors and interventions as causally relevant variables, do they share the same commitments, and if not, which one of them figures in the elaboration of invariance under intervention? Finally, what is the principled relation between invariance and interventions, does the former incorporate the latter in its characterization, or does the latter extend the former?

7.1 The formal characterization of interventions – the logic of interventionism meets the theoretical features of interventions.

As to characterizing what interventions are we will utilize, and in turn expand the formal characterization that was presented in chapter 1.
As you might recall, we gave the following characterization of an intervention in the first chapter: an exogenous variable that "fixes" the value of a selected variable (X), while "switching off" the influences of other variables (Z) on X relative to a second variable Y, such that there is a causal chain running from I to Y through X (Woodward 2003, page 98).
Before heading to characterizing the formal definition(s) of interventions, we will make the distinction between two senses of intervention, the notion of intervention variable (for X with respect to Y), a type-level formulation of an intervention, and the notion of intervention (on X with respect to Y), the token-level formulation of an intervention.
When discussing interventions, we will first start with presenting the characteristics of an intervention variable (IV), and then move on to characterizing token interventions (IN), subsequently returning to focus on (IV) throughout the rest of the chapter, as that is the formulation of interventions that comes the closest to testing interventions, a third formulation of interventions that are used to test for invariance under intervention.

Woodward presents the following four conditions that an intervention variable, I, has to satisfy if it is to qualify as an intervention variable for X with respect to Y. – Instead of paraphrasing the conditions, I'll rather give a verbatim presentation of them, as what we aim at discussing is not the conditions themselves, but rather their presuppositions and implications.

"(IV)

11. I causes X
12. I acts as a switch for all the other variables that cause X. that is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.
13. Any directed path from I to Y goes through X. That is, I does not directly cause Y and it is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the I-X-Y connection itself: that is, except for (a)

any causes of Y that are effects of X (i.e., variables that are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.

14. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X" – Woodward 2003, page 98

Similarly for the definition of an intervention (given the notion of an intervention variable).

(IN): "I's assuming some value $I = z_i$, is an intervention on X with respect to Y if and only if I is an intervention variable for X with respect to Y and $I = z_i$ is an actual cause of the value taken by X" (Woodward 2003, page 98).

One of the apparent and most striking features of IV is the way it implicitly invokes invariance conditions into the characterization of what it is to be both an intervention variable and intervention for/on X with respect to Y.
What I have in mind is an idea that we came across earlier, namely modularity, the feature that a causal system (of causal relationships) exhibits when it allows us to intervene on a variable without disrupting the causal structure in the system as a whole. – Formulated in formal terms, what we aim at capturing when requiring that interventions preserve the causal structure of the system being intervened on, is that the intervention preserves the arrows (graphical language for directed causal pathways) that are directed out of a variable being intervened on, while simultaneously breaking other arrows that are directed into the variable being intervened on (Woodward 2003, page 101).
This feature does not only concern the variable X, being intervened and the other variables Z, being held fixed/influenced by the intervention on X, but it generalizes to the structure as a whole in such a manner that, "all other arrows, both those directed out of the variable intervened on and arrows directed out of an into other variables are preserved intact" (Woodward 2003, page 102). Understood this way, it is built into the idea of an intervention and its graphical representation that it changes some things and leaves other things unchanged.

The reason this is feature is especially important, is that it reveals that an interventionist account of causation commits itself to an analysis, interpretation and representation of causal relationships that integrates invariance conditions into the causal system under investigation. In practice, this amounts to the presupposition that we specify the boundary conditions of the

causal system under investigation, the variables and their range of values, the dependency relations among the variables, while in turn distinguishing between those aspects of the system that remains unchanged from those that change under interventions[13].

In addition to the modularity/invariance under intervention, interventions (IV + IN) exhibit the following features:

Relativity: the notion of intervention (IV + IN) on X is defined only relative to a second variable Y, which bars the existence of an intervention on X simpliciter. – This feature is an extension of the role interventions figure in as causal assessors. In order to an Intervention I to assess if, how, and to what extent X causes Y, then this requires information about both the causal relationship between the two variables, but also information about the causal structure running from I → X → Y (as in the definition of IV) (Woodward 2003, page 103).

Nonanthropomorphism: the conditions laid down in IV and IN does not make any references to human activities or what human beings can or can't do, which renders it fully conceptually and logically possible that an intervention satisfies IV and/or IN despite being practically impossible to carry out for a human being. – What the conditions do make reference to, are causal and noncausal terms such as "cause" and (statistical) independence (how these terms are used will be discussed in due time) (Woodward 2003, page 103).
This feature is an extension of the objectivity that is incorporated into Woodward's formulation of interventionism, which stands in contrast to alternative formulations of interventionism that makes extensive use of human activity/agency as means of defining interventions, thereby rendering them subjective in nature (example: Menzies and Price, 1993). (Woodward 2003, page 103)

Noncircular and nonreductive causal information: the causal information that is required to characterize the notion of intervention on X with respect to Y consists of the four following pieces of information; "(i) information about the causal relationship between the intervention

[13] What I will argue is that this distinction is reminiscent of the familiar distinction between boundary conditions and laws that figures in most, if not all, formulations of physical theories. The latter specifies the aspects of an representations of reality that is subject to changing values and conditions, while the latter specifies those aspects that remains invariant under transformation of boundary conditions.
The reason this familiar resemblance is worth emphasizing, is that the distinction in physical theories concerns ontology and facts about the world, and the intimate relationship between invariance and physical reality, which in turn gives us a specific case study of how an ontological formulation of the same interventionist distinction might be given. – See Wigner for a classic account of the distinction between boundary conditions and laws.

variable I and X, (ii) information about whether there are other causes of Y that are correlated with I, (iii) information about whether there is a causal route from I to Y that does not go through X, and, (iv) information about the counterfactual claim concerning the behavior of Y under interventions on X" (Woodward 2003, page 104-105). – The causal information *is circular* in the sense that utilize a conception of "cause" (and related causal terms) in order to assess the causal relationship between X and Y with the help of the intervention variable I, *but it is not viciously circular* in the sense that we need to know beforehand whether X is a cause of Y.

This is connected to the idea that interventionism is nonreductive in the sense that causal information and assessment is centered around a 'circle of concepts' that are either implicitly or explicitly causal, and that there is no ambition (nor possibility) of reducing these concepts to non-causal concepts (Woodward 2003, page 20, 104)

Putting these two features together, what the fundamental idea underlying the interventionist analysis of causation aims at capturing is that "(…) we can explain what it is for a relationship X and Y to be causal by appealing to facts about other causal relationships (or the absence of such relationships) involving I, X, and Y and to counterfactual claims concerning the behavior Y under interventions on X." (Woodward 2003, page 104-105)

Before turning to discussing one of the two central topics of this chapter, the causal relata of interventionism, there is a further point concerning the causal information that should be brought to attention, the role and presuppositions of non-causal correlations in the interventionist framework.

Woodward embraces the widespread opinion that causal inference on the basis of correlational evidence alone is not possible. This is a consequence of the underdetermination of correlational evidence, a predicament that is based on different causal structures being compatible with a given body of correlation evidence (Woodward 2003, page 106). As a way of singling out, and in turn determining which of these structures that are correct, Woodward points out that, we must make use of additional background information; "either domain-specific about whether certain variables do or do not cause others (…) or domain-general principles, such as the Causal Markov Condition connecting causal and correlational information or both" (Woodward 2003, page 106).

This additional information, which is a part of the two interventionist formulations of causation, M and TC, establishes a connection between causal and correlational claims.

Not only does these additional hint at causal structure incorporating content that goes "over

and above the claim that certain correlations obtain", but they also show that the manner in which differences in causal structures are reflected in differences in patterns of correlations, is an empirical question that is subject to whether interventions (or related possibilities) happen to occur. (Woodward 2003, page 107). Most importantly, there is no conceptual guarantee that it is possible that causal structures will be reflected in patterns of correlation. (Woodward 2003, page 107).

The remarks concerning the presuppositions and role of patterns of correlation within the interventionist framework is worth pondering upon for a number of reasons.
(i) The distinction between differences in causal structures and differences in patterns of correlations, seem to presuppose a distinction between 'potential' and 'actual' differentiation, where the former follows from the counterfactual commitments of interventions (particularly, the result of different combinations of interventions). While the latter, follows from the non-causal (hence non-modal?) correlations that hold as an empirical fact, with variables taking certain values depending on the circumstances.

(ii) The distinction between, on the one hand, "empirical question", the idea that whether there is a correlation between causal structure(s) and patterns of correlation(s) is sensitive to whether interventions (or other sorts of occurrences that are equivalent to interventions) happen to occur. And, one the other hand, "conceptual guarantee", the idea that interventions (or other sorts of occurrences that are equivalent) as an extension of their meaning (and regardless of whether interventions or any other equivalent process occur), informs about a correlation between causal structure(s) and patterns of correlation(s), seems to strengthen the argument from the preceding chapter that the semantics of interventionism has an empirical dimension. However, it also seems to exclude that this empirical dimension is built into the semantics of interventionism (like the distinction between causal generalizations/truth-conditions and causal singularities/evidential conditions from last chapter). – This distinction, will prove to be important when analyzing the metaphysics of interventionism, as it will be necessary to distinguish between treating metaphysics as built into interventionism, and metaphysics as following from interventionism.

(iii) Finally, the idea that the transition from 'empirical question' to 'conceptual guarantee' depends on whether or not interventions occur (in addition to whether causal-correlational patterns are semantic or not), seems to establish an intervention-to-world relationship, where the truth-value of causal-correlational claims are sensitive to objective considerations (as in

how the world actually is and how the pattern of correlation reflects the state of the world). – This is an important affirmation of the empirical dimensions of interventionism, which is connected to the metaphysics of interventionism.

These three points illustrates that the logic of interventionism exhibits a complex structure that balances between epistemological, ontological and methodological considerations, and where answers to questions concerning topics such as metaphysical commitments, traditionally takes an absolute 'either-or' form, should rather be thought of as relative 'if-so' answers that admits of degrees and gradations along the different dimensions (ontology, epistemology and methodology). – This is a topic that we will return back to when discussing invariance under intervention in detail, as the shift from thinking in absolute to gradual terms is a central feature of invariance.

7.2 Interventions, variables, and causation – variables as causal relata and representational devices.

This brings us to one of the fundamental aspects of interventionism, the topic of causal relata. So far we have framed the discussion about interventionism in terms of variables, while talking interchangeably about them as representations of causal relationships and relata figuring in causal relationships.
The clearest examples of the way variables have figured extensively in the characterization of interventionism, is through the idea that causal relationships are understood as the relation between variables, and that dependency relations among variables are defined in terms of how change in the value of a variable X influences the value of another variable Y (Woodward 2003, page 39, 45).
In this section, we want to clarify the distinction between variables as representations of causal relationships and variables as causal relata, with the aim of discussing the assumptions and implications of relating causation to variables.

The preceding characterization of interventions embodies this idea by requiring that an intervention on X satisfies the following criteria; (i) there must be a unit or entity that is characterized by X, (ii) there must be a range of (at least two) values that it is possible for X to possess, (iii) there must be a well-defined notion of changing the value of X (for whatever unit or entity is characterized by X), (iv) the intervention must cause the same unit or entity to

possess a different value of X (Woodward 2003, page 111).
The reference to there being a 'well-defined' notion of changing the values of variables, seems intuitive at the outset and in light of the guiding idea that causal relations are relations that can be used for manipulation and control. However, as will be highlighted, the notion of 'well-defined changes in the values of variables', imposes a number of restrictions on both the interventionist analysis and the admissible causal relata, restrictions that we will now turn to highlighting.

First, the choice of selecting variables as causal relata is not an arbitrary, rather, it follows from the logic of interventionism that presupposes that if something is to qualify as a cause, then it must be possible to describe what it would be like to change or manipulate it (Woodward 2003, page 112). This suggests that the natural way to think of causal relationships within an interventionist framework, is to think of them as relations among variables, or more precisely, "(…) as changes in the values of variables, where one of the characteristics of a variable is that it is capable of taking two or more values and of being changed from one of these values to another." (Woodward 2003, page 112).

The standard causal relata for type-causal claims in philosophy are properties or event types, and it is when trying to translate talk of causal relationships between properties and event types into variable talk that the full force of the requirement 'well-defined' changes of values are felt.
An example of how to perform such a translation is to express the occurrence or nonoccurrence of the properties or event types in question in terms of 'indicator variables' that takes the value 1 or 0, depending on whether the properties or event types occurs or not (Woodward 2003, page 112).
The cases we are interested in, are not those where such translations are feasible, but rather, those cases where such translations are unfeasible.
The cases I have in mind, are those in which "claims about relationships between properties and event types are not readily understandable as claims about relationships between variables because the whole idea of changing those properties does not seem to be well-defined." (Woodward 2003, page 112).

The reason these cases are particularly interesting are *not because of the fact that* they fail to satisfy the requirement of being well-defined, but rather *because of the reason* that they fail to be well-defined. What I have in mind are the three examples of characteristics of

properties/event types that Woodward refers to as preventers of translation. The three preventers may take the following forms: logical, conceptual and/or metaphysical considerations (Woodward 2003, page 112).
There are (at least) three ways in which these three considerations prevent translation of properties/event types to variables; (i) by being properties that necessarily must be possessed by every object and which can only take one value (for example: objects being present), (ii) by exhibiting properties that we lack a well-defined notion of changing (for example: if an object can travel faster than the speed of light), (iii) by being properties that we lack a coherent conception of their alternative states (or conjunction of states) (for example: what it is to exist but to be non-physical). (Woodward 2003, page 112-113). – The idea is that if properties exhibit characteristics that for logical, conceptual or metaphysical reasons fail to translate properly to variables, then this indicates that the former is not well-defined, which results in the properties/event types under question being inadmissible as causal relata in an interventionist framework. The converse of this restriction, is that the interventionist framework imposes a set of requirements, ranging from conceptual to metaphysical, that the causal relata figuring in the causal claims and relationships have to satisfy (and in fact does satisfy) if they are to figure in the interventionist framework.

However, Woodward notes that failure of well-definedness for properties does not prevent claims or generalizations that contain them from being true (as they might reflect facts about the world), rather, these claims and generalizations are ruled out when it comes to qualifying as causal claims/relationships.
There are numerous implications following from the requirement of properties/event types being well-defined, and they may be grouped together in the following manner: there are those implications that concerns the admissibility of properties/event types as causal relata. Given the requirement of well-defined properties/event types, there will be a number of properties that traditionally have been thought to be causal, which in an interventionist framework will be disqualified as causes. Examples of such properties include “the property of being member of a certain species, being a member of a particular race, and being a certain age”.

Furthermore, there are those implications that concerns the response to the unclarity found in the inadmissible properties. For some properties, the source of unclarity results from a similar unclarity concerning “what hypothetical experiments to associate with the properties in

question", which in practice translates to the properties being unmanipulable.
As a way of removing the unclarity surrounding these properties, clarifying the meaning of the causal claims they are embedded within, and thereby make the claims and properties amendable to the interventionist analysis, Woodward proposes that we can solve the issue by "replacing them with claims involving variables, the values of which are manipulable, and specifying more precisely just what it is that they imply about the outcomes of hypothetical experiments (Woodward 2003, page 114). – Hence, well-definedness implies that it must be possible to replace unclear properties with variables consisting of a range of (at least two) values, and finally, relate these values to the outcome of hypothetical experiments if we are to translate unmanipulable properties into variables/causal relata within the interventionist framework.

Finally, there are those implications that concerns the response to the presence of a multiplicity of possible interpretations (and the multiple hypothetical experiments associated with them) of the cause variable under which it is manipulable. In these cases, the solution to clarifying the meaning of causal claims involving unmanipulable and/or unclear properties, lays in the possibility of stating explicitly, "which hypothetical experiment is the intended interpretation of a causal claim (…)" (Woodward 2003, page 114). – The implication of well-definedness in these situations, is that it is possible to explicate and demarcate an intended interpretation of hypothetical experiment, such that it maps into a causal claim, which in turn clarifies the meaning of the causal claim under question.

In a summarized fashion, what well-definedness implies for the causal relata figuring in the interventionist framework, is that the properties are manipulable, translatable to variables (with a range of values and associated hypothetical experiments), and that there is an explicit connection between the hypothetical experiments and meaning of causal claims. – These and the other requirements figuring in the characterization of interventions (IN and IV), are closely connected to the role of M, TC, and IN/IV as regulative ideals – characterizations providing information about what conditions a causal relations must fulfill in order to be true (or in other words, what ideals they must fulfill in order to be true) (Woodward 2003, page 114).
This is an important feature that further affirms the prescriptive and hypothetical nature of the interventionist framework. What matters for the assessment of causation is not (only) whether it is true that the relationships we encounter satisfy the ideals and requirements we stipulate

for a relationship to be causal (descriptive), rather, it is equally important to investigate what would have to be true about those relationships for them to satisfy those ideals and requirements (prescriptive).

7.3 The metaphysics of interventions – when objectivity meets modality to yield a realist thesis of causation.

We have now come to the central aspect of interventions, metaphysics, or to be more precise, the metaphysical commitments of interventions. In section 1.6, we touched upon some of these commitments when discussing truth-conditions, reproducibility and the evidential basis of counterfactual claims. The idea was that the interventionist framework presupposes the existence of possible (hypothetical) interventions, mind-independent dependency relations (among variables/causal relationships), and objective truth-values of causal claims. (Woodward 2003, page 45, 118, 119)

Each of these commitments falls under the heading of 'objectivity of causation', or as Woodward describes it, 'Realism about Causation'.

In this section, we want to expound these commitments by unpacking the 'realism' that underlies the interventionist analysis of causation, highlight how this relates to the metaphysics of interventionism, and further, clarify the stance Woodward takes on metaphysics in general (as a way of preparing the way for discussing invariance and its metaphysical commitments).

As a starter, we present some of the background for the discussion. Woodward's realist understanding of causation stands in (or is at least meant to) in contrast with so-called anti-realist (or subjectivist) accounts of causation. These accounts are characterized by two shared distinctions within the components constituting causal relationships, and a shared assumption about the nature of causal relationships.

The two components; (i) *objective entities and/or relationships* that are *"out there" in nature*, and that are normally characterized in terms of its commitment "the existence of certain regularities or correlations" and *facts* about "spatiotemporal relationships" (Woodward 2003, page 118). (ii) *subjective facts* that are somehow *constructed or introduced by us*, having to do with *our psychology or mental organizations* (for example; our expectations concerning the persistence of certain regularities and/or our practice of organizing our experience in certain ways (ibid).

As to the shared assumption; the anti-realist/subjectivist accounts commits themselves to the idea that the difference between causal and non-causal regularities is characterized in terms of a difference in "our beliefs or attitudes regarding these two classes of regularities", rather than an intrinsic and objective difference (located in the relationships themselves) (Woodward 2003, page 118). – It is particularly this latter feature, in addition to the (ii) component, that underlies the idea that subjectivist accounts reduce causation to "a "projection" onto the world of our experience of human agency and that causation is thus a "secondary quality"" (the reference to color expresses that, just as color needs to be understood relative to a perceiver's standpoint, causation needs to be understood relative to an agents standpoint as well), as Menzies and Price (1993) (classic example of a subjectivist and manipulationist account of causation) would put it.

This is the background against which Woodward formulates his alternative account, a realist/objectivist account of causation, an approach that we now turn to characterizing.
First, how does Woodward's realist account relate to the subjectivist 'reduction' of causation to projection and belief states/attitudes?
Quick answer: skeptical towards the idea of projection due to unclarity of what 'projection' means, but open to the possibility that subjective considerations influences conception of causation (albeit in a more limited manner than the subjectivist).
The detailed answer: as the most straightforward reading of the subjectivist 'reduction' translates into relating the truth-value of causal claims to the existence of human being (with their beliefs, attitudes and experiences), this reading would render the truth-values of causal claim dependent on the existence of human beings (and where variation in the former would be subject to variation in the latter). – Woodward contends that this might be true in a limited sense, as our assessment of which causal claims we accept as true will be influenced by which possibilities we consider as serious (as discussed earlier, serious possibilities are sensitive to both subjective as well as objective considerations).
However, and as we discussed in detail earlier, once these serious possibilities are fixed, there is no further sense in which the counterfactual conditionals are sensitive to, or dependent on human attitudes and beliefs (provided that the causal claims in question do not concern the effects or causes of psychological states). – What this gives us is a sketch of the realist underpinnings of Woodward's account, particularly, the commitment to mind-independent truth-values for counterfactual conditionals:

Mind-independence of truth-values for counterfactual conditionals: "(…) The counterfactuals on which causal claims are based seem to be true or false in a mind-independent way, even if it is true that the causal claims themselves reflect additional assumptions about which possibilities are serious" (Woodward 2003, page 118).

As a way of furthering the realist underpinnings of Woodward's account, we might turn the attention to the structural relationship between interventions, invariance and mind-independence.
Assume that we start from the subjectivist point of departure of relating causation to an agent's beliefs and actions, imagine a situation where an agent wonders whether she can change Y indirectly through X (in other words, change Y by changing X). In this situation, there will be some sense in which it is up to the agent (in other words, dependent on her beliefs and actions) whether she chooses to bring about X, but regardless of this choice, it is a presupposition of her deliberation that it is not also up to her whether X occurs, hence if Y occurs. Unpacked further, this presupposition implies; "that if it is possible to change Y by intervening on X, then there must exist an independent invariant relationship between X and Y that the agent makes use of when she changes X, and in doing so, changes Y" (Woodward 2003, page 119). – The idea underlying this presupposition is that despite starting from a subjectivist point of departure, and analyzing causation in terms of means-ends, objectivist commitments will find its way into the analysis, as the process of deliberation and act of intervention presupposes an independent, invariant, and existing relationship between X and Y, if the former is to figure as a mean to change the latter. – The conclusion holds for Woodward's own formulation of hypothetical interventions that do not make any references to human agency or physical possibility as well.

This gives us a second constituent of the mind-independence that figures in the realist/objectivist account of causation, and that may be divided into two parts:

Mind-independence of hypothetical interventions: if a causal claim, X causes Y, is true, then there is a causal relationship between the causally relevant variables (X, Y). What it is for the causal claim to be true, is that there exists some possible intervention, I (on X), that would make it possible to change Y by intervening on X, hence rendering the causal claim, X causes Y true.
Contraposing this idea, while combining it with the mind-independence of causal relationships (and truth-values for counterfactuals), establishes the presupposition that there

exist mind-independent hypothetical interventions (in virtue of the truth of the above causal claim). (Woodward 2003, page 45, 59, 65).
This form of independence is built into the interventionist conception of causation, and it follows from the interaction between the role of interventions, the modality of interventions and the structure of interventionist causal relationships.

Mind-independence of causal relationships/dependency relations: "(…) if it is possible to change Y by intervening on X, then there must be an independently existing, invariant relationship between X and Y that the agent makes use of when she changes X and, in doing so, changes Y – a relationship that would exist and have whatever characteristics it has even if the agent were unable to manipulate X or did not exist. (Woodward 2003, page 119).
This form of independence is particularly important due to its central role in the interventionist framework. Woodward points out in a three-step argument (3.3.1-3.3.3), how the existence of mind-independent (invariant) causal relationships is built into the manipulationist conception of causation, in the way interventions, invariance and independence mutually substantiate each other both empirically and theoretically (Woodward 2003, page 120).

The presence of mind-independence and its relation to an agent's deliberation and actions, seems to highlight a general point about realism, objectivity and mind-independence. Each of these terms locates the assessment, domain of application and manner of variation within the dimension of the world, which is another way of stating that they concern ontology/metaphysics.
In other words, for something be objective and mind-independent, and hence be part of a realist understanding of causation, is for there to be something (this being an ontological entity, property and/or relation) in the world that renders it true that this is the case (as subjective considerations are by ruled out by definition).

The preceding points provides an overview of the constituents that constitutes the 'realism' that Woodward commits himself to. An equally important point, is how this realism relates to the metaphysics of interventionism, particularly, what kind of metaphysical commitments this introduces beyond the existence of mind-independent (invariant) causal relationships, hypothetical interventions, and objective truth-values for counterfactual conditionals?

Woodward answers this question negatively, or at least, ambivalent in the sense that he points out quite explicitly "I emphasize that the kind of realism that follows from this way of viewing matters is metaphysically modest and noncommittal It requires only that there be facts of the matter, independent of facts about human abilities and psychology , about which counterfactual claims about the outcome of hypothetical experiments are true or false and about whether a correlation between C and E reflects a causal relationship between C and E or not. Beyond this, it commits us to no particular metaphysical picture of the "truth makers" of causal claims." (Woodward 2003, page 122).

In this passage, Woodward seems to restrict the realist underpinnings of interventionism to objective truth-values for counterfactual claims (outcome of hypothetical experiments that are true or false independently of facts about human abilities and psychology) and objective dependency relations between variables figuring in correlations (facts about demarcation between causal and non-causal correlations that are independent of facts about human abilities and psychology).
Finally, Woodward denies explicitly that realism does not commit us to any 'metaphysical picture of "truth-makers", which is a considerable claim, given the centrality of truth-conditions within the interventionist framework.

The clue to whether this passage reveals all there is to the metaphysics of interventionism and its realist underpinnings (which at first glance seems to leave us with a deflationary view of metaphysics), lays in unpacking the constituents of the 'facts' about counterfactual claims and correlations, while tying together the fragments of presuppositions, commitments remarks that might reminisce so-called 'truth-makers' for causal claims.

As we discussed earlier (see section 1.5-1.7), part of what constitutes the fact that truth-values and dependency relations are objective, is that they are mind-independent.
When it comes to truth-values, mind-independency implies that there exists a possible hypothetical intervention that would make the causal relationship, X causes Y, figuring in the counterfactual claim under question come out as true (and where the existence of this hypothetical intervention and question of whether the claim would come out as true depends on the world/causal structure (after we have fixed which possibilities to take seriously)). (Woodward 2003, page 45)
As for the dependency relations/correlations, mind-independency implies that the causal correlations are realized in token correlations with the help of actual interventions (evidence

that the causal correlation holds), and further, that whether these dependency relations hold is sensitive to how the world/causal structure is (after we have fixed variables to include in the causal model). (Woodward 2003 page 119)

These consideration does not contradict Woodward's claim that his view is 'metaphysically modest' (what this is supposed to mean is left undefined for now), but they seem to the claim that his view is 'noncommital' under question, as the considerations that figures into 'facts' do not only impose commitments on our representation of causal relationships, but also commitments about the causal relationships themselves and their relation to the world/structure they are embedded in. – Much more will be said about the details and extent of these commitments, but for now it suffices to note that there is more to them than Woodward admits in the above passage.

Before turning to discussing the claim about the metaphysical picture of 'truth-makers', it should be noted that Woodward 2007, develops his view on truth-makers in a number of ways, where the most considerable difference is revealed in the approach to the notion, where Woodward 2003 takes a more quietist approach (does not elaborate on the details of what truth-makers are and what their role are, but makes indirect remarks about it), Woodward 2007, takes a more sympathetic approach (recognizes and discusses the place and role of truth-makers in the semantics of interventionism, but with qualifications, in accord with the ambitions of metaphysical modesty).

Another development consists of extending the adherence to the idea of truth-makers, to introducing an explanatory relationship between truth-makers and truth-conditions (the traditional candidate for this relationship, 'grounding' has often had a metaphysical slur), and furthermore, to also extending the idea that truth-makes consists of a range of factors (that goes beyond facts about spatiotemporal and physical considerations).

As Woodward notes concerning the idea that true causal (and counterfactual) claims requires some sort of grounding (or 'truth-makers') in fundamental physical laws: (…) This idea strikes me as arguably correct if it is interpreted in the following way: given a true garden variety causal claim, there will be some associated in – principle physical explanation (or story or account, to use more neutral words) for its holding, and this will include, among other factors, appeal to fundamental laws" (Woodward 2007, 103-104).

Now that this remark is in place, I return to discussing the metaphysical picture of truth-makers. In discussing the truth-conditions for causal claims, a number of requirements were

put in place to ensure that the meaning and interpretation of the causal claims (and their corresponding hypothetical interventions) where in accord with the generality (causal generalizations/type-causal claims), clarity (well-defined counterfactual conditionals), and objectivity (mind-independent truth-values/causal relationships) of the interventionist analysis of causation.
These truth-conditions, and their related the truth-makers (conditions under which the causal claim comes out as true), specifies the content of these truth-makers; the existence of possible (hypothetical) interventions, the existence of a reproducible causal generalization that satisfies an invariance condition (remains invariant under intervention), and a difference-making/change-relating relation between cause and effect.

Putting these observations together yields a condition that the metaphysical picture that the truth-makers reflects must satisfy the above requirements, while at the same time balance between the empirical and hypothetical aspects of the interventionist framework. – The metaphysical picture of an interventionist truth-maker, will thereby have to incorporate a number of features while simultaneously performing a number of tasks.

The question is whether this picture takes a specific form, and what this tells us about the metaphysics of interventionism?
I contend that a viable candidate for an interventionist truth-maker, is the hypothetical intervention (in all its senses), and where invariance under intervention is the central feature of truth-conditions that specifies its application, character and domain.
However, it should be noted that causal correctness follows from a causal claim/relationship remaining invariant under intervention, and where interventions are metrics for assessing the conditions under which this is the case. – what follows from this claim, is that highlighting the metaphysical picture of a truth-maker in an interventionist framework requires that we investigate the metaphysics of both hypothetical interventions and invariance under interventions.

As a way of substantiating this claim, while highlighting what hypothetical interventions and invariance under intervention might tell us about the metaphysical picture of truth-makers, I propose that we turn to the task of investigating the modality of interventions.
Put in other words, investigating in what sense interventions must be possible, and what this tells us about the metaphysical picture of interventionist truth-makers.

7.4 Interventions, modality and metaphysics – clarifying the metaphysical picture of truth-makers and truth-conditions

An important feature of hypothetical interventions that has so far been treated loosely and intuitively is the manner in which it is 'possible' to intervene on X, if X is to cause Y.
We touched upon the requirement of non-anthropomorphism, which rules out human and/or technological capabilities as candidates for the type of possibility that figures hypothetical interventions. We also touched upon the requirement of well-definedness, which rules out candidates that for conceptual, logical and/or metaphysical reason fail to be translated into talk of variables (and change of values of variables).

Putting these two requirements together yields the following (preliminary) characterization of a sense in which interventions on X must be 'possible' if X is to cause Y: "(…) interventions on X must at least be logically possible and well-defined." (Woodward 2003, page 128).

An important question is what the relationship between the above sense of logical possibility relates to one of the traditional candidates for a sense of possibility, namely physical (nomological) possibility (possibility that have to do with consistency with the laws of nature and initial conditions)
There are two readings of physical possibility, a strong notion and a weak notion of physical possibility:

Strong notion of physical possibility: "On one notion, and event E is physically possible if and only if it is consistent with the laws of nature and the actually obtaining initial conditions. When conjoined with determinism, this notion of physical possibility implies that interventions on X will not be possible unless they actually occur (…)". (Woodward 2003, page 128). – This sense of possibility is ruled out due to being too strong. This sense of physical possibility implies that an intervention has to occur if a causal relationship is to hold (which stands in contrast to the hypothetical commitments of the manipulability theory that allows a causal claim/relationship to hold regardless of whether an intervention occurs).

Weak notion of physical possibility: "E is physically possible if and only if there is some set of possible initial conditions ("possible" in the sense that the conditions themselves are consistent with the laws of nature), perhaps different from those that actually obtain, such that

the occurrence of E is consistent with those conditions and the laws of nature. In other words, E is physically impossible if and only if its occurrence is ruled out by the laws of nature alone, independently of facts about initial conditions." (Woodward 2003, page 128) – This sense of physical possibility is also ruled out due to the possibility that a causal claim X causes Y is true, but where an intervention on X is not physically possible (in the weak sense).

An example of this is sense of physical possibility, is when a causal claim comes out as true, but where an intervention on the cause variable is not physically possible as it fails to satisfy the conditions for an intervention, IN as a matter of physical fact[14].
(A concrete example of the previous claim: "changes in the position of the moon with respect to the earth and corresponding changes in the gravitational attraction exerted by the moon on various points on the earth's surface cause changes in the motion of the tides", which translates to the counterfactual claim that 'if the gravitational attraction exerted by the moon is varied by varying its distance from the earth, the motion of the tides would change'. (Woodward 2003, page 129)).

A second example of this sense of physical possibility, is a causal claim where "C's only occur spontaneously in the sense that they themselves have no causes", but which rules out the possibility of physically intervening on whether C occurs, but which allows for C to have further effects E (Woodward 2003, page 130). – Conjoining the underlying idea in this example with the requirement of physical possibility yields a third notion of physical possibility:

Revised weak notion of physical possibility: If C is to cause E, it must be physically possible to intervene on C, which translates into the requirement that if C is to have effects, it must itself be such that it can be affected by some other cause (Woodward 2003, page 130). This notion of physical possibility introduces a second consideration in addition to the causal relationship between C and E, namely, the causal history of C. The idea is that if C is to figure as the cause of E, then C needs to have a physically possible causal history (ibid).

[14] (As when an intervention by definition are required to be fine-grained surgical processes, but where a cause as a matter of physical fact, fails satisfy this requirement due to being too coarse-grained. A difference in grain corresponds to whether a causal chain I → C → E can be established without having additional effects E_2, and where fine-grained surgical operations establishes the above chain, whereas coarse-grained operations fails to do so).

Woodward rules out this sense of physical possibility on the ground that it conflicts with the idea that the truth-value of a causal claim (should) depend on the nature of the relationship between C and E, and not on whether C is caused (or could be). This conclusion is reinforced by the fact that whether the causal history of C is a consideration that is extrinsic and irrelevant to the nature of the relation between C and E (Woodward 2003, page 130).
This brings up an important feature of the interventionist analysis of causation that has been understated so far, namely, the possibility of making a logical and conceptual distinction between, on the one hand, the causal connection between C and E, and, on the other hand, the causal history of C (including the causal connection between an intervention I, other factors Z, and C). – I touched upon a similar distinction when discussing the difference between the internal/external validity of causal claims, and where it was noted that the invariance conditions (including reproducibility, scope and range) of a causal claim/relationship hinges on the distinction between the background conditions specifying the domain of invariance and the causal relationships that are embedded within these conditions. (Woodward 2003, page 71)

In light of Woodward's denial of the three above senses of physical possibility, where does this leave the role and requirement of physical possibility?
In light of the underlying idea that IN represents a regulative ideal that characterizes the notion of an ideal experimental manipulation (rather than an actual manipulation), and relatedly, what would happen under hypothetical interventions (rather than what happens under actual interventions), physical possibility does not seem to play a role in the characterization of interventions, nor does it seem to be required that interventions are physically possible. - As Woodward himself puts it: "(…) as long as there is some basis for assessing the truth of counterfactual claims concerning what would happen if various interventions were to occur it doesn't matter that it may not be physically possible for those interventions to occur." (Woodward 2003, page 130).

The idea of restricting, and in turn, eliminating the requirement of physical possibility on the grounds that it is not needed for 'assessing the truth of counterfactual claims' is driven by a means-ends methodology where the motivation for introducing and constructing an analysis around the notion of interventions (ends) guides the features that we include in the characterization of interventions (means).
The treatment of physical possibility is an illustrative exemplification of this methodology.

The idea is that we start with a set of ends consisting of a list of requirements, match these ends with a set of 'appropriate' means, assess whether and to what extent the means fulfills the ends in terms of whether they succeed. – The following argument illustrates the point:

Ends: "We want to exclude cases in which we had no coherent conception of what it was to change the variable intervened on" (Woodward 2003, page 130)

Ends: "We wanted to exclude cases involving confounding – cases in which, although an association between C and E persists under changes in C, this is due to something other than a direct causal link from C to E, such as an independent cause of E that is correlated with C or a direct effect of the intervention itself on E." (Woodward 2003, page 130)

means: "The notion of an intervention was designed to ensure that if the changes in C had a certain kind of history, these sorts of possibilities would not arise." (Woodward 2003, page 130-131)

Requirement: "This [the means] suggests that there will be a basis for claims about what will happen to E under an intervention on C *as long as we can associate some well-defined notion of change with C* and *as long as we have some grounds for saying what the effect, if any, on E would be of changing just C and nothing else*." (Woodward 2003, page 131; my own emphasis)

Requirement: "There must be *a way of disentangling – perhaps merely conceptually or analytically rather than in actuality* – the effect on E of changing just C from the effects on E of changes in other potentially confounding variables, including direct effects from the intervention process itself." (Woodward 2003, page 131; my own emphasis)

Conclusion: Provided that disentangling the effect on E of changing C, from the effects on E of changes in other potentially confounding variables is possible with the help of logical/conceptual possibility alone, then this provides reasons (on methodological grounds) for not requiring physical possibility.

The means-ends argument establishes that physical possibility does not figure in the modal basis of hypothetical interventions, that consists of counterfactual, logical and conceptual possibility.

The argument is substantiated by the following facts;

(i) The fact that the means does not require reference to physical possibility (provided that the

above requirements are satisfied).
(ii) The fact that the modalities that do figure in the basis/disentanglement (counterfactual, conceptual, and logical possibility) might fail to be physically possible (yet satisfy the conditions for an intervention on X with respect to Y to occur).
(iii) the fact that there will be some causal claims that are associated with a notion of intervention that fails to correspond with a physical process, due to the latter being too coarse-grained (hence failing to satisfy IN) (Woodward 2003, page 130-133).

More generally, the above considerations illustrate that counterfactual claims and possibilities does not reduce to nor supervene on physical claims and possibilities, which makes counterfactual possibility an independent modality within the interventionist framework. – Relating these observations to the discussion about the metaphysical picture of interventionist truth-makers (with the hypothetical intervention being the selected candidate for a truth-maker) yields three positive points and three negative points concerning the 'nature' of truth-makers.

Positive points: (i) truth-makers must be modal in nature, (ii) truth-makers must be logically possible and well-defined (in an interventionist sense of relating to variables), (iii) truth-makers must be objective in the sense of existing mind-independently.
Negative points: (i) truth-makers does not presuppose spatiotemporal continuity nor possibility, (ii) truth-makers cannot be separated fully from subjective considerations (variable choice, serious possibilities), (iii) truth-makers are not reducible to non-causal and/or non-modal facts, relationships and/or terms.

Comparing the positive and negative points might give the impression that spatiotemporal considerations and possibilities are disconnected from the counterfactual basis and possibility that figures within the interventionist analysis of causation.
The reality is that spatiotemporal considerations constitutes an integral part of interventionism, with the clearest examples being; first, that causal generalizations relate spatiotemporally located individuals and that change-relating describes changes that are located in these particular individuals (Woodward, 2003, page 40).
Secondly, that spatiotemporally located physical interventions figure in the evidential basis for counterfactual claims (Woodward 2003, page 72-73).
Thirdly, that spatiotemporally located causal relationships exist independently of and prior to intervention (Woodward 2003, page 119-120).

In light of these considerations, it is important to distinguish the role of spatiotemporal continuity and physical possibility in the characterization of the modal character of hypothetical interventions, from the role of spatiotemporal individuals, relationships and interventions in the characterization of the semantics of interventionism.

Given that counterfactual conditionals lays at the heart of the interventionist framework, the logic of interventionism will be influenced by the considerations that figure in the modality (See for example contrastive focus and the structure of causal claims/relationships on p. 146)[15].

7.5 Contrastive focus – causal structure, representation and modality meets invariance conditions

I conclude this chapter by noting an important implication of centering modality at the heart of interventionism, which serves as an example of how the nature of (counterfactual) modality influences the logic of interventionism, and relatedly, the nature of invariance under intervention.

The implication I have in mind is contrastive focus.

In chapter 2, I discussed the details of the truth-conditions causal claims, with particular emphasis on the logic of interventionism (including its presuppositions, suppositions and implications).

This was cashed out in terms of hypothetical interventions, counterfactual dependence, and invariance under intervention.

[15] This aspect of the structure of the content of causal claims is quite central as from an epistemological point of view, it suggests that we need to possess counterfactual information about the different states of a system that we are trying to describe causally, while at the same time convey information about the range/conditions of the correlation between the variables in a system of interest. While, from a metaphysical point of view, contrastive focus introduces the question of addressing what it is for a system to be in different states. Given that we want to understand what change from one "state" to another involves, we need to disentangle (or at least clarify) what it is for a system being in different states from what it is for the variables that represent the states of the system to take different values.
I concede that this is a tricky problem. In some cases it makes sense to interpret this relationship interchangeably, such that talk of different values and different states are interchangeable salva veritae. In other cases (and I think these represent the majority of the cases we are interested in) where we are interested in evaluating a range of different models in light of their ability to explain and represent the causal structure in a physical system, we need to distinguish the different states of the system from the different values of the variables.

The underlying idea; if there is some possible intervention on X that would change Y under the right circumstances, then the causal claim, X causes Y would be true. The right circumstances describe the circumstances under which the causal relationship figuring in the causal claim would continue to hold under intervention, that is, circumstances under which it would remain invariant under intervention.
As a way of conveying information about the truth-conditions of causal claims and the dependency relations among the variables figuring in the causal relationships, Woodward introduces two devices, mathematical/functional relationships and contrastive focus (Woodward 2003, page 66-67).
There are two senses in which contrastive focus occurs in the interventionist framework, as a representational device, and, as a feature of causal structures.

As for the first sense, contrastive focus (in an interventionist framework) consists of the following components; a range of situations $\{S_1, S_2, \ldots S_n\}$, a range of possible values $\{v_1, v_2, \ldots v_n\}$, a set of variables $\{V_1, V_2, \ldots V_n\}$, a specified situation $\{S_@\}$, and an actual value $\{v_@\}$ (taken by a specified variable $\{V_n\}$ in the sense that the value is mapped into the variable, which indicates that every value is a property of a variable)[16]. – The logic of contrastive focus provides metrics for assessing patterns of dependence among a set of variables $\{V_1, V_2, \ldots V_n\}$, by contrasting the actual value of the variable(s) $(v_@)$, in a specified situation $(S_@)$, with the possible values in other situations $((\text{f}: \{v_1, v_2, \ldots v_n\} \rightarrow \{V_1, V_2, \ldots V_n\}) \subset \{S_1, S_2, \ldots S_n\}, (where\ S_@ \neq S_n)$ in order to explain the conditions under which the value $\{v_@\}$ rather than $\{v_n\}$ holds in the specified situation $\{S_@\}$.

The idea is that we can specify information about the range of situations under which the patterns of dependence would hold by using the specified situation as a background condition against which the actual value functions as a metric for contrasting the patterns of dependence in the specified situation with possible situations and values (Woodward 2003, page 67-68).
Contrastive focus is clearly modal in the sense that it does not restrict the assessment of patterns of dependence to the specified situations and actual values, but makes references to

[16] This point concerning the mapping of a value to a variable, a point that applies generally, is a theoretic choice of making an implicit presupposition explicit, by noting how there is no such thing as 'free-floating' values within an interventionist framework, as values are properties of variables. – In light of this assumption, it would have sufficed to refer to values, while omitting the reference to variables (as it is implicit that the existence of values presupposes variables), but I have chosen to include both values and variables for clarity.

possible situations and values, which are then integrated into the basis for assessment of patterns of dependence.

This is the first sense of contrastive focus, a device for conveying information about patterns of dependence and boundary conditions[17].

As for the second sense of contrastive focus, the logic remains the same, but its application shifts from representation to description of dependency relations.
We can think of the above model, as describing the structure of causal claims, the set of variables $\{V_1, V_2, \ldots\ V_n\}$ (which includes a cause variable C, effect variable E, and intervention variable I), as describing causal relata (in a causal relationship), the set of values $\{v_1, v_2, \ldots\ v_n\}$, as describing different states of the variables, and finally, the situations $\{S_1, S_2, \ldots\ S_n\}$ as describing the dependency relations between the different states of the variables, with each situation describing the contrast between intervening on, and in turn changing the cause C from one state $C_@$ to an alternative state C_n, with the putative effect E, changing accordingly from state $E_@$ to E_n (as before, $S_@$ figures as the background against which we contrast the different states of the variables against, with $(S_@ \neq\ S_n)$).

The idea is that it follows from the interventionist analysis of causation, and, is built into the content of causal claims that conveying information about what will happen under hypothetical interventions, requires conveying information that one or more specific change in the cause will change the effect or the probability of the effect. – As Woodward notes "This in turn means that all causal claims must be interpretable as having a contrastive structure,

[17] It should be noted that the usage of contrastive focus was popularized by Mackie (1974) and his popular "chestnut example", in which he proves how a counterfactual analysis of causation requires more specific information beyond 'If X, then Y', if it is to yield a correct verdict of counterfactual conditionals. – This specific information is conveyed with the help of restricting the assessment of causal relationships to "(…) a context, against a background which includes the assumption of some causal field." (Mackie 1974, page 34-35). Applying the idea of a causal field to the topic of causal analysis includes translating, and in turn expanding causal questions such as ' What caused this explosion?' into "(…) What made the difference between those times, or those cases, within a certain range, in which no such explosion occurred, and this case in which an explosion did occur?" (Mackie 1974, page 35-35). – Jonathan Schaffer (2005) illustrates and generalizes these ideas to yield a comprehensive account of contrastive causation, in which Woodward's interventionist conception of causation is an instantiation of.
The specific information is captured by contrastive focus, and the introduction of this representational device is substantiated by the idea that counterfactual conditionals hold under some, but not all conditions, which requires that we specify the boundary conditions of the counterfactual conditionals under question.

and it also has the implication (…) that to causally explain an outcome is always to explain why it rather than some alternative occurred" (Woodward 2003, page 146).

This is the second sense of contrastive focus, a feature of the causal structure that figures in causal claims[18].

The existence of patterns of dependence are demonstrated by the situations under which the patterns would hold, while their characteristics are described in terms of the conditions (including range, scope and depth of dependence among the variables/values) of the situations that the patterns hold under.
Contrastive focus describes both the existence and characteristics of patterns of dependence by highlighting the situations under which they would hold, the manner in which they would hold, and to what extent they would continue doing with the help of an implicit (correlation between the values x and y, rather than x* and y*), and/or, explicit (' the occurrence of C rather than not-C caused Y, causes Y rather than not-Y) rather-than locution.

If some pattern of dependence exists, there is some (range of) situation(s) $S_@$, some under which it would hold, rather than some alternative (range of) situation(s) S_n ($Where\ S_@ \neq S_n$). Understood in this way, the (range of) situation(s) $\{S_@, S_1, \dots\ S_n\}$ under which the patterns of dependence would hold determines the characteristics of the patterns of dependence by conveying information about how and to what extent the changes and differences in the set of variables $\{V_1, V_2, \dots\ V_n\}$, and their corresponding values $\{v_1, v_2, \dots\ v_n\}$, would be correlated with each other (with the I, C and E, specifying the direction of change-relating/difference-making). – Hence, contrastive focus provides means for expressing the boundary conditions of patterns of dependence that were discussed in the

[18]Kenneth Waters provides an analysis of contrastive focus, difference-making and causation that elaborates on the details of contrastive focus and the facts it aims at capturing. By developing the logic of contrastive focus, Water introduces further metrics for assessing the different ways and manners in which contrastive focus captures facts about change-relating and difference-making.
Waters introduces the distinction between actual difference making and potential difference making causes, the former designating the cause that actually made a difference, in contrast to the latter that that designates the cause(s) that would make a difference (the central feat of potential difference making is that causes that qualify, will also qualify as causes in an interventionist sense (hypothetical interventions)). (Waters 2007, page 568).
Relativized to the current discussion, the notion of actual difference making corresponds to the pattern of dependence that figures in the specified situation $S_@$ with the specified value $v_@$, while the notion of potential difference making corresponds to the range of situations $\{S_1, S_2, \dots\ S_n\}$ and values $\{v_1, v_2, \dots\ v_n\}$.
Water's discussion is illustrative for a number of reasons, which I will return to when discussing invariance, but for now it suffices to note that Waters highlights the importance of specifying the values, variables and situations that we use as background conditions for assessing the truth of causal claims.

last chapter under the heading of reproducibility, truth-conditions and boundary conditions (as mentioned earlier, and as will be illustrated more precisely later, this hints at the importance of invariance under intervention within the interventionist framework).

Understood in an interventionist framework, the rather-than locutions are accompanied by an explanation of why the patterns hold in some situations and not others.
The explanation: that the patterns of dependence remain invariant under intervention in some situations, but fails to do so in other situations. – If the patterns of dependence remain invariant under intervention in some situations S, rather than some other situation S*, then this explains why the patterns of dependence hold in S rather than S*.
Hence, contrastive introduces a two-tiered sense of counterfactual modality, (i) counterfactual modality that concerns what would happen under hypothetical interventions, and, (ii) counterfactual modality that concerns what would happen under/in different conditions/situations[19].

This concludes the discussion of contrastive focus and the semantics of interventions in general. As have been hinted to numerous times throughout these chapters, I propose that the discussion of the logic, characteristics and metaphysics of interventionism can be tied to a discussion of invariance under intervention in the following manner.

The proposal is that the features of the truth-conditions of causal claims that have been discussed so far, the existence of hypothetical interventions, causal generalizations, reproducibility and counterfactual conditionals (just to name a few), may be thought of as a subset of invariance conditions.

As have been pointed out several times, truth-conditions of causal claims invokes invariance conditions both implicitly (as in the emphasis of type-causal claims and causal generalizations), and explicitly (as in the invocation of reproducibility and modality), which makes invariance conditions suitable candidates for assessing the truth-conditions of causal claims. – Understood this way, discussing the truth-conditions of interventionist causal claims

[19] This distinction is further expounded by Woodward (2003, page 248) when he discusses the different forms of changes that are relevant for assessing the invariance of causal generalizations, with counterfactual dependency (i) corresponding to invariance under interventions, and counterfactual dependency (ii) corresponding to invariance under non-interventions (such as changes in background conditions of the generalizations). – More will be said about this.

and relationships is a matter of discovering and assessing their invariance conditions. As Woodward himself points out quite explicitly; "(…) invariance is the key feature a relationship must possess if it is to count as causal or explanatory" (Woodward 2003, page 239).

My goal for discussing the logic, characteristics and metaphysics of interventionism, has been to prepare the way for discussing invariance under intervention, by describing the interventionist framework that invariance conditions are intimately connected to and influenced by. – With this background information in place, I now turn to focusing solely on invariance under intervention by ending with two questions: what exactly is invariance under intervention, and how does invariance conditions relate to metaphysics?

8 Invariance under intervention – the mark of causation and gist of explanation

8.1 The logic of invariance under intervention – unpacking the connection between invariance under intervention and the methodology of interventionism

The debates of the philosophy of causation concerns a wide array of topics and questions. In metaphysics, discussions often concern what we take causation to be and how it relates to the laws of nature, in epistemology; how we acquire knowledge about causal claims and how it relates to statistical data, and in methodology; what role modelling and pragmatic considerations play in the study of causation. – Interventionism touches upon and addresses a number of topics ranging from explanation and modelling to metaphysics and epistemology. Interventionism as a theory contributes to the development of an account of causal assessment, the study (or 'problem' as Russo 2014 calls it) of establishing of what causes what, while interventionism as a framework, allows us to articulate a methodology for causal assessment, which in the literature is known as causal modelling.

Causal assessment, as seen through the lens of interventionism consists of testing for whether a correlation between two variables, X and Y, is invariant under intervention. Invariance under intervention is a necessary (and sufficient) condition for a correlation to be causal in the sense that; "(variable) X causes (variable) Y if, and only if, were we to manipulate X, Y would accordingly change, and the relation between the two would remain stable, or invariant, under a sufficiently large class of interventions or manipulations of the putative cause variable" (Russo 2014, page 158).

This idea generalizes naturally to interventionism as a whole, as Woodward points out when opening the chapter in *Making Things Happen* that explores invariance: "the guiding idea is that invariance is the key feature a relationship must possess if it is to count as causal or explanatory. Intuitively, an invariant relationship remains stable or unchanged as various other changes occur" (Woodward 2003, page 239)

This gives us a superficial picture of the relation between causal assessment and invariance

under intervention. – What we will be concerned with in the remaining thesis, is unpacking and characterizing what "invariance under intervention" consists of, what role it serves in the study of causation (with some remarks addressing causal explanation), and finally what metaphysical (if any) the feature has on the logic of interventionism.

In characterizing and unpacking what invariance under intervention consists of, I will distinguish between two domains and three strands in the study of invariance. The domains demarcates the two topics that Woodward's interventionist framework emphasizes in the study of causal explanation (where invariance plays a central role).
The two domains: (i) the methodology of causal modelling (in particular, how we test for invariance), and, (ii) the semantics of causal claims (in particular, what it means for a causal relationship to be invariant).
The three strands marks the central topics that we will study in detail when developing an account of invariance under intervention.
The three strands: (i') the structure of invariance (what invariance consists of/how we represent and define it), (ii') the metaphysics of invariance (what invariance implies/requires of the world), and (iii') the epistemology of invariance (how we discover/understand invariance).

When characterizing and discussing invariance I will restrict my attention to invariance as a qualitative notion, while at most making remarks about the quantitative form of invariance. – The motivation behind this choice is that the considerable features of invariance as described and discussed in the characterization of the qualitative aspect, whereas the quantitative aspects merely fill in the details of how to actually use the invariance to characterize specific relationships in terms of their exact values.

As a way of addressing each of these features of invariance under intervention, I will draw contrasts between the logic of invariance under transformation as discussed by Nozick (2001) and Woodward (2003, 2007) due to the structural similarities of the two, and the illustrative example that the former presents by connecting the discussion of invariance to objectivity.

8.2 The logic of invariance II – Characterizing the dimensions and metrics of invariance under transformation

The strategy Nozick (2001) pursues, starts out with characterizing the notion of objectivity, apply it to facts/truths, thereby defining the notion of an objective fact, and then, drawing the connection between objective facts and invariance under transformation by illustrating how the latter term (invariance under transformation) underlies and explains the features that figures in the former term (objective facts). (Nozick 2001, 75-76).

After having established this connection, Nozick unpacks and elaborates upon the features that underlies the logic of invariance under transformation; admissible transformations [specified changes that figures as metrics for assessing objectivity/invariance] (Nozick 2001, page 79), incompleteness of objectivity [invariance conditions as transmitters rather than generators of objectivity] (Nozick 2001, page 79), necessary truths versus deep truths [range/scope of invariance across possible worlds versus degree of invariance under specified transformations] (Nozick 2001, 83-84), ordering of objectiveness [objectivity/invariance as gradual rather than dichotomous] (Nozick 2001, page 87), gradation of invariance [range of invariance as a metric for degree of objectivity] (Nozick 2001, page 87), assessment of objectivity [actual score along some dimension of transformations/occurent objectivity versus potential score along some dimension of transformations/underlying objectiveness] (Nozick 2001, page 88), and finally, dimensions of truth [that truth admits of degrees along the orderings; objective versus subjective, absolute versus relative] (Nozick, 2001, page 101).

As a way of motivating the identification of degree of objectivity with range of invariance, Nozick points to the fact that invariance conditions and symmetries (physical equivalent of invariance) figures extensively in physical laws and generalizations (Nozick 2001, page 81). – To name some examples; requirement of covariant formulations of laws in general relativity, group theory in quantum mechanics, symmetries n particle physics, and finally, Emmy Noether's conservation laws (that establishes a connection between satisfaction of invariance conditions and conservation of some quantity in the universe) (ibid).

Rather than contrasting every feature of Nozick's characterization of invariance under transformations, I will restrict myself to assessing the idea of admissible transformations as

metrics for invariance and the ordering of objectiveness as a gradual, rather than dichotomous notion. – As a way of illustrating the utility of these two features for illustrating the logic of invariance, I will draw upon the conception of truth that follows from conjoining invariance and objectivity; that truth admits of degrees, and that gradation admits of relativizing to some dimension/ordering (Nozick 2001, page 101).

Provided that invariance under admissible transformations admits of degrees, and that Nozick's account of the objectivity of facts and truths incorporates invariance conditions, this implies that variation in the set of admissible transformations admits of degrees (Nozick 2001, page 100).
The two orderings of truth in Nozick's framework is the dichotomies totally objective versus extremely subjective and "wholly absolute" versus "extremely relative" (that despite representing dichotomies, admits of degrees).

Incorporating, and in turn, applying the logic of invariance under transformation to the assessment of truth, leads to the following conception of truth:

1. **Claim:** "Any given truth can be located at a position along each of the two above orderings" (Nozick 2001, page 101). → As a way of establishing this claim, hence constructing a logic for assessing the objectivity/truth of some fact with the help of invariance under transformation, Nozick idealizes the assessment and location of truths in a three-step process.
2. **Step 1:** "Imagine that each ordering is a complete ordering" (ibid). (That is, when we specify the ordering we utilize to assess the admissible transformations, we suppose that the ordering is exhaustive)
3. **Step 2**: "Suppose that numerical value(s) van be given to a facts/truths position along each ordering" (ibid).
4. **Result**: Each ordering will qualify as a dimension.
5. **Step 3**: "Suppose that each ordering forms the axes of a two-dimensional space" (ibid).
6. **Result:** "Each truth, then, possess a position within this two-dimensional space" (ibid).

The central aim of characterizing the logic of invariance under transformation with emphasis on its application to truth, is that it maps into Woodward's own formulation of invariance conditions, truth-conditions and its relation to metaphysics.

Woodward's own treatment of interventionism and invariance conditions, when discussing the metaphysics of interventionism, cashes this out in terms of 'modest realism', the thesis that there are objective (mind-independent) "facts of the matter (…) about which counterfactual claims about the outcome of hypothetical experiments are true." (Woodward 2003, page 121).

If Woodward's own formulation of the logic of invariance under intervention embodies the above implications for a conception of truth, this provides grounds assessing truth in terms of invariance conditions.
If invariance conditions admit of degrees, then so will truth-conditions. If the dimensions that figures in the above assessment of truth-conditions are sensitive to how the world is (in virtue of being objective and mind-independent), this provides grounds for assessing the degree of metaphysical implications in terms of truth-conditions, which in turn is a matter of assessing invariance conditions (as illustrated earlier in chapter 2, when discussing the role between invariance- and truth-conditions).

Before making this inference and establishing the connection between Nozick's logic of invariance under transformations and Woodward's invariance under intervention, I will provide turn to characterizing the central features underlying invariance under intervention.

8.3 Interventionism and invariance under interventions – definition, logic and metaphysics

1. A property that applies to generalizations (where a generalizations is one that is invariant under certain changes) [2003:239]
2. A property that is gradual in both its domain and scope [2003:240]
3. A property that involves both a threshold and continuum (part of the relativity of invariance). – Necessary for a generalization explanatory/causal generalization to be invariant under at least some interventions, but possibility of distinguishing between different generalizations with respect to *the range or kind of interventions* and *other*

sorts of changes under which they are invariant. → it is necessary to specify under which changes the relation is (or is not) invariant to acquire a complete picture of this threshold, continuum and contrastive evaluation.

   a. "Different sort of causal claims will be associated with different possibilities regarding the changes under which those relationships may be stable that are relevant to assessing invariance. → Different sorts of causal claims will be associated with different sorts of claims about which relationships are invariant and about the changes under which they are invariant". [2003:244] → "For the most part, my focus in this chapter is on issues having to do with the invariance of specific functional relationships".
   b. Further elaboration on the threshold and continuum: A necessary condition for a generalization to qualify as invariant at all amounts to invariance under at least some testing intervention, while among generalizations meeting this condition, there will be differences in range of invariance (in terms of differences in background conditions and non-I-changes) over which the generalization would continue to hold [2003:254]. – See [2003:257] for yet another elaboration on threshold and continuum.

4. Appeals to change-relating and modality in the sense that a generalization is explanatory in virtue of answering w-questions, while the notion of change-relating requires invariance under certain changes and interventions. [2003:247]
5. **Different forms of changes distinguishes what kind of invariance conditions a generalization satisfies [2003:248].** – There are three types: (i) invariance under changes in background conditions to the generalization. (ii) invariance under changes in those variables that figure explicitly in the generalization itself; two subcategories of such changes (a) changes that result from an intervention on variables figuring in the generalization, (b) changes that does not result from an intervention (non-i-changes)

   The central form of invariance that matters for the assessment of explanatory status: invariance under intervention on the variables figuring in the generalization [2003:249]
6. **How to assess and test for invariance (correctly predicting the value of the dependent variables in the generalization by intervening on the independent variables):** in order to test for whether a generalization is invariant, then we apply a

testing intervention. – A testing intervention does not merely change the value of X, but it changes it in such a way that according to generalization G, the value of Y will change under the intervention.

   a. Woodward gives the following characterization of what it means for a generalization G to be invariant under a testing intervention: "*G is invariant under this testing intervention if and only if it correctly describes what the new value of Y,y_1 would be under this change: that is, if and only if it remains true that G (x_1) = y_1 for the system S. Invariance under at least one testing intervention (on variables figuring in the generalization) is necessary and sufficient for a generalization to represent a causal relationship or to figure in explanation." [2003:250]*
   b. Testing intervention as an actual physical change in the value of X for the system S + the invariance of G as a result of a counterfactual (rather than factual) intervention → Invariance under testing intervention as a modal notion that applies to physical systems [2003:25], which is one of the clearest realist commitments of Woodward's project. (see summary for remarks concerning this point)

7. Invariance is a property that has limited range/scope in the sense that the generalization will remain invariant under a certain range of invariance, while the assessment of invariance is always relativized to a particular system [2003:251]. → **These two features has the following implications for the explanatory import of invariant generalization:** (i) a generalization will be explanatory with respect to some aspect of the behavior of a system as long as it is invariant under some (appropriate) interventions. (ii) What matters for explanation is invariance of a generalization with respect to the particular system X whose behavior we are trying to explain. [ibid]
8. **Connecting between interventionism and invariance**: The manipulationist account of causation is intimately intertwined with invariance under testing interventions for causal generalizations (in other words, interventionism amounts to the adherence of invariance condition on change-relating generalization) [2003:253]
9. **Connection between invariance and causation/explanation:** Invariance under testing interventions amounts to a necessary and sufficient condition for a generalization to represent a causal/explanatory relationship. [2003:253]

10. **Connection between invariance and explanatory depth**: Connection between Invariance under testing interventions amounts to a connection between range of invariance and explanatory depth; invariance under larger and more important set of changes corresponds to increased explanatory depth. [2003:257]
11. **Invariance domain**: The range of invariance under interventions that described a relationship that can be exploited for purposes of manipulation of control. This amounts to the idea of tracking a pattern of counterfactual dependencies, which according to Woodward is at the heart of successful explanation [2003:259]
    a. **How to characterize and asses the size of invariance domains (sets and subsets of changes and interventions as basis for assessment).** "*To a very good degree of approximation, the range of changes and interventions over which (6.4.1) is invariant is a proper subset of the range of changes and interventions over which the generalizations (6.4.2.) of the deeper engineering theory of the behavior of the car are invariant. That is, any change that will disrupt the latter will also disrupt (6.4.1.), but not vice versa. Thus, any properly behaved measure will assign a larger size to the domain of invariance of the latter*." [2003: 260]
    b. **The assessment of invariance as a partial ordering:** considering that not all generalizations may be structured into a set-subset relation, and that there is nothing about the size of the set of changes that automatically yield any importance or privilege, then we get at most an ordinal comparison (numerical illustration of invariance).
    c. **Assessment of invariance as a matter of emphasizing privileged changes and explanatory status:** The idea is that depending on the subject, context and domain, then there will be some set of privileged or particularly important sorts of changes that matter to the assessment of invariance. – The idea is that depending on our explanatory interests, there will be some changes that have a fundamental explanatory status (regardless of the size of the domain).
    d. **The ground for assessing invariance as a subject and domain specific notion**: Expectations about the sorts of changes over which fundamental relationships will be invariant help set the explanatory agenda. However, despite the seemingly pragmatic feature of these expectations, then Woodward emphasizes that they will be grounded in "very general empirical discoveries

about the sorts of relationships in the domain of these disciplines that have been found to be invariant in the past and under what sorts of changes, which in turn invites to interpreting expectations as "objective". [2003:263] **(see notes in the summary concerning this section)**

12. **Concluding remarks concerning the assessment and construction of explanatory generalizations:** "Comparisons of invariance based on the proper subset relation and judgments about the significance or importance of the intervention over which a generalization is invariant – play an important role in the construction and assessment of explanatory generalizations" – together, they provide a partial basis for distinguishing among invariant generalizations with respect to degree and kind of invariance and for judging that, although a generalization is invariant under some interventions, it is nonetheless relatively fragile or unrobust in the sense that it is stable only under unimportant set of interventions or under a set of changes that is relatively small in comparison with some rival generalization". [2003:265]
13. **Connection between invariance and lawhood:** According to Woodward, it is the range of interventions and other changes over which a generalization is invariant and not the traditional criteria that are crucial both to whether it is a law and to its explanatory status (it is only the support of counterfactuals that plays a role). – This is the basis for regarding laws (as commonly understood) as just one kind of invariant generalization. [2003:267]
14. **Connection between invariance and scope:** "In general, scope differs from invariance in at least two ways. First, invariance is a modal notion: it has to do with whether a relationship would remain stable under various hypothetical changes. In contrast, scope, as I understand it, is an "actualist" nonmodal notion: t has to do with how many systems or how many different kinds of systems there actually are for which a generalization holds in the sense described above. Second, insofar as the notion of invariance applies to change-relating generalizations it requires stability under intervention." [2003:270]
15. **Connection between assessment of invariance and counterfactuals:** Woodward distinguishes between on the one hand, "other object" counterfactuals that "describe what the behavior of objects other than o would be under the counterfactual circumstances in which they are A". While, on the other hand, "same object" counterfactuals describe how the very object o would behave under an intervention. –

Woodward emphasizes and adheres to "same object" counterfactuals when appealing to modality when assessing the range of invariance and the explanatory status of generalizations [2003:281].

The choice and preference of the form of counterfactuals is not arbitrary. Considering that explanation in Woodward's framework is a matter of tracking what an outcome depends upon and that this is done with the help of interventions, then "same object" counterfactuals shares many of the structural features of Woodward's conception of explanation. In addition to the structural similarities, "same object" counterfactuals involving interventions have the advantage of being "clear enough in meaning", and are often accompanied by the ability or potential of obtaining "scientific evidence that is relevant to their truth" [2003:282].

16. **Testing for invariance in light of counterfactual reasoning in order to yield an invariance space.** When assessing the counterfactual support for a generalization, the invariance-based approach suggests that we expand the basis for assessment in such a way that we do not restrict ourselves to one set of interventions or changes when assessing generalizations such as (6.9.1) [All As are Bs]. Rather, the invariance-based approach suggest that "we think in terms of a whole family of counterfactuals whose antecedents correspond to different interventions or different changes in background circumstances under which A would be true and then ask whether B would also be true in those circumstances". – The idea is that depending on the truth-value of the whole family of counterfactuals, then we get a more thorough and detailed description of the range of "circumstances or interventions over which the generalization is invariant". [2003:284]. This provides us with the ability to construct an invariance space that contains a number of different ways for the world to be when assessing the truth-value of the antecedent A.

    Traditionally in a Lewisian analysis of possible worlds, we restrict our attention to a single world or set of worlds we deem "closest" to the actual world, which (almost) by default restrict the assessment to the world(s) in which A is true, and thereby ask if B is true in one (or all) of the worlds. Woodward's proposed analysis of possible worlds differs from this strategy by expanding the set of worlds such that we may "regard it as legitimate to consider counterfactuals in which A occurs under conditions that are very dissimilar from those that hold in the actual world" [2003:284].

    Woodward elaborates on the details of this revision of the counterfactual analysis in

such a way that it not only allow us to handle generalizations of the form (6.9.1), but also generalization such as (6.9.2) [If x were to be an A, then it would B]. – By pursuing the same strategy, we are able to expand the basis of assessment to a whole family of counterfactuals "corresponding to different ways of strengthening the antecedent of (6.9.2).

The result is the construction and consideration of a set of counterfactuals of the form "(6.9.2*): "If A and C were true, then B would still be true", where "A" is brought about by different possible interventions and "C" represent different (possible) background circumstances in which A might occur, and depending on the truth-value of the different combinations of A and C, then the counterfactual will be assessed differently.[20].

17. **The difference between paradigmatic laws and invariant generalization**: (i) invariant generalizations have narrower scope (ii) invariant generalizations will break down under a wider range of "extreme" value + possible changes in the background conditions. [2003:285]
18. **The difference between different types of invariant generalizations in terms of interfering conditions**: "In general, the set of possible "interfering conditions" for (6.6.3) is very large and heterogeneous and will resist any simple, informative characterization. (…)By way of contrast, although paradigmatic laws like Maxwell's equations do break down under certain extreme values of the variables figuring in those equations ,whether the equations hold or not depends just on the values of those

[20] This way of drawing the distinction between "possible" interventions/background circumstances and (hypothetical) interventions/background circumstances as a characterization of the range of invariance space, reminds of Waters (207) distinction between potential difference making causes and actual difference making causes.
If we abstract away from Waters own formulations and rather use the current terminology, then possible difference making causes corresponds to the "possible interventions" that depending on which "possible" intervention is assessed/held fixed (A*), then the background circumstances (C) of how A is brought about will be influenced. While the actual difference maker will correspond to the (hypothetical) intervention that we are currently strengthening (A) (and thereby holding fixed will assessing different counterfactuals). – It is the actual difference making cause that determines the outcome and occurrence of A, but the possible difference making causes (A*)plays a part in the assessment of the range of invariance (as members of the whole set of counterfactuals).
This comparison invites further questions, due to the fact that Waters (2007) regards this distinction as "ontological". Depending on the viability of this comparison, then it might commit Woodward to a similar distinction (or at least a story) when describing the difference between members of the "set of counterfactuals" and "possible interventions" (given that they serve the same function as Waters possible/actual difference making causes).

variables and not on how those values are brought about". [2003:285] (see comments concerning this section for further important remarks)

19. **The difference between different types of invariant generalizations in terms of background conditions:** "When the circumstances which paradigmatic laws fail to be invariant are known, they typically can be given a relatively simple, unified characterization. Such circumstances seem to fall into one of two categories: laws break down either for extreme values of variables that explicitly figure in them (e.g. high temperatures and pressures, in the case of the ideal gas law) or when some very small set of variables that have been omitted from the law diverge from a limiting value – the pattern being that the law holds when the variables take this limiting value but not otherwise". [2003:286]
20. **The difference between paradigmatic laws and invariant generalizations as a matter of degree rather than kind:** The difference between the two forms of invariant generalizations seems to amount to a difference in the scope of their domain of invariance, range of interventions, and changes in background conditions over which these generalizations are invariant. – Paradigmatic laws are characterized by their simplicity, wide scope and invariance under large and important set of changes that can be given clear characterization. Whether we regard a generalization as a law depends on whether it exhibits some (or all of) these features we regard.
21. **The difference between Woodward's invariance based account of laws and the MRL theory:** The MRL theory seems to emphasize the number of instances, number of different kinds of instances and the subsumption of different regularities of a contingent generalization that is regarded as a law of nature. – Woodward's invariance based account does not emphasize any of these features when accounting for the explanatory status of generalizations (scope is the closest analogue to the emphasis on instances, but this must be understood differently than in the MRL theory). [2003:289] Furthermore, the MRL theory emphasizes "what actually happens or, to be more precise, to which Humean regularities are realized and to how frequently they are realized" when accounting for what the laws in a specified domain of investigation (world(s)) is. While the invariance-based approach emphasizes what would happen in different counterfactual circumstances, in such a way that our expectations for paradigmatic laws is that they continue to hold when "various changes in initial conditions and interventions occur" [2003:289].

The central contrast between invariance and Skyrms notion of resiliency: Resiliency concerns the epistemology of an agent's belief, where resilient beliefs would have a stable subjective probability when conditionalising on other truth-functional beliefs in some family. Invariance is an objective notion in the sense that it "has to do with the way the world is, rather than with beliefs. *Invariance as Woodward sees it, concerns the extent and accuracy of the descriptive power of a generalization when it comes to describing the behavior of some system (or the relationship described by the generalization) "under changes that are actual physical manipulations or alterations in the system.* [2003:299-300]

8.3.1 Revised characterization of invariance under intervention

The strength of the causal relation provided by the notion of total cause without (without including invariance) [connection between intervention and causation]. The interventionist analysis of causation does not provide any more information than the fact that for the relation between X and Y to be causal, all that is required is that there is a (single) intervention on X that would be associated with a change in Y [2007:76]

Invariance as a necessary and sufficient condition for a generalization to describe a causal relationship/distinction between invariance and stability: For a generalization to qualify as causal, then it is both necessary and sufficient that it is invariant under at least some interventions. – Furthermore, causal generalizations will differ in the invariance conditions that they hold under, where some of the conditions corresponds to changes that apply to variables that figure in the causal generalization, while other conditions corresponds to variables that does not figure in the causal generalizations. The former set of conditions are those that are referred to and termed "invariance under interventions", while the latter is referred to (and this is a revision of the original proposal) as stability under background conditions. – In Woodward (2003), both of these type of condition was labelled as a form of invariance, whereas in this paper distinguishes between invariance and stability, where invariance has to do with the conditions that corresponds to interventions in the variables that figure in the generalizations, thereby being crucial for assessing the causal status of the generalization in question. [2007:77]

The relativity of invariance and stability as and the limit of invariance: Invariance and stability are relative in the sense that a generalization can be invariant under some range of interventions and not under other and similarly for stability under background conditions. –

This to the idea that in order to explain the fact that a generalization holds and why generalizations differ in when it comes to size of the domain of invariance, then it is necessary to include other conditions and factors that contributes to the invariance/stability of the generalization in question. [2007:77]
This feature leads us to the new feature that Woodward incorporates into his analysis of invariance, namely incompleteness.

Invariance and incompleteness – elaborating on the range and form of limitation:
Woodward gives the following example of an incomplete causal generalization that has a limited range of invariance (the example of the spring is a paradigmatic example that is utilized in Woodward (2003) to characterize invariance) "*the restoring force exerted by a spring is contingent on many additional conditions besides those specified in (H) – conditions having to do both with the internal structure of the spring and with its environment. (…) In this sense, (H) describes a causal relationship that is merely locally invariant and stable or a relationship of contingent or conditional dependency."* [2007:78].
– Incompleteness in this sense seems to correspond to the idea that the invariance/stability of a causal generalization depends on external conditions (where external refers to conditions that are not specified nor part of the content of the variables figuring in the generalizations), thereby rendering it contingent. Combined with the limited invariance mentioned earlier, the causal generalization will be locally invariant in the sense of holding under a limited range of changes conditional on exhibiting invariance/stability under a range of contingent conditions. The strength and size of the domain of invariance will correspond to the range of dependence in both the changes figuring in the variables and the changes in the conditions of the generalization.
Local invariance characterized more formally: *"One of the simplest possibilities, illustrated by the case of the spring, is that the value of one variable Y depends on the value of a second variable X according to some relationship Y = G (X) when some third variable B assumes some value or range of values* $B = [b_1 \dots b_n]$ *but this dependence disappears or changes radically in form when B assumes other values outside this range" [2007:78].* – The crucial idea in this section is a two-tier dependency, where the first characterizes the range of invariance between the variables X and Y, while the second characterizes the range of invariance between the functional relationships between "Y = G (X)" and "$B = [b_1 \dots b_n]$". Provided that it is a "contingent matter of fact" that B usually takes the specified values in the domain of investigation and that an intervention on X does not disrupt the functional

relationship "Y = G (X)" by changing the value of B to a value outside of its range of invariance, then the generalization in question will be locally invariant.

The two-tiered account of dependency and range of invariance is one of the most important and considerable revisions of the original account, as it seems to reinforce the connection between range of invariance, state of affairs (understood as facts about the environment) and the objectivity of causation. – Woodward reiterates this point by stating the following concerning the connection between contingent facts occurring in causal generalizations in the special sciences and the objectivity of causation.

"Whether occurrences like meteor strikes are likely to occur around here is, of course, a contingent matter, dependent not just on the fundamental laws governing nature but on initial and boundary conditions that might have been otherwise.[21] (…) Note though that although contingent, it is an 'objective' matter (not a matter of idiosyncratic individual taste and opinion) whether the sorts of disrupters of (S) just described are likely to occur around here" [2007:80][22].

[21] The idea that there are initial and boundary conditions that function as truth-conditions for causal claims (together with fundamental laws governing nature/physical explanation), but that the former are not reducible nor subject to the latter seems to raise further question as to how Woodward's interventionist theory of causation relates to the idea of causal foundationalism. – In light of the idea that there are distinctions and considerations that matters for the evaluation of the truth of causal claims (that are not governed by laws) and that these are objective (as in not subject to our interests), we need to clarify the following issue. In order to satisfy the preceding considerations while enabling assessment of causal claims, we need to determine whether and how 'contingent facts' relate to the 'world' and how this is independent from physical laws. – See Woodward (2007: p. 70) for comment on causal foundationalism, the discussion between Ney and Glymour about causal foundationalism for more background information, and Firsch (2014) for a response to Woodward's (2007) paper.

[22] Invoking «contingent facts» as a way of specifying and assessing the range of invariance for causal generalizations is reminiscent of the discussion in Woodward (2003) about "serious possibility" as a way of discriminating between the variables and conditions that we should include in our causal models when assessing the truth of various counterfactuals. – An interesting and important discussion concerns the extent and nature between the assessment of "serious possibility" as a pragmatic (hence, non-objective in the sense of not being mind-independent) endeavor and the assessment of contingent facts as objective (in the sense of mind-independency).

Woodward himself remarks in a footnote in Woodward (2007: page 80) that it is a difficult questions as there are differing pragmatic and theoretical considerations that pull in opposite directions. – I will return to this issue when discussing the metaphysics of invariance, and in particular, how the objectivity of causation relates to the subjectivity of variable choice.

See: Woodard (2007: footnote, page 90) for further remarks concerning the relation between the objective and subjective features of his framework.

9 Conclusion

In the introduction I noted how Woodward's metaphysical commitments and presuppositions follows from the role it serves within his interventionist framework, and the logic of interventionism.

What I hope to have established is that the metaphysics of invariance under intervention follows from; (1) the role it fulfills within the interventionist framework, which in summarized fashion, relates to the way causal assessment, interventions and causal correctness presupposes invariance conditions (and that these conditions have metaphysical commitments insofar as the realist underpinnings of interventionism are present).

(2) That the logic of interventionism, in virtue of incorporating and implying invariance conditions introduces metaphysical presuppositions and commitments that are a function of the range, scope and depth of invariance (as illustrated on both methodological grounds as in means-ends metaphysics and on metaphysical grounds as in the 'modest realism).
In the second chapter we noted how the methodological considerations that underlies Woodward's characterization of interventionism both constrains and guides the metaphysical consideration that figures within our theories and accounts of causation.

In the third chapter, we characterized some of the central features of interventionism, with particular focus on the intersection between interventions, invariance and metaphysics. – The idea that facilitated this intersection, was the conjunction of the existence of hypothetical interventions, presupposition of invariant causal generalizations, and the mind-independence and objectivity of both preceding conjuncts (including the objectivity truth-values for the results of the combination of interventions).
In the fourth chapter, we discussed the logic, definition, role and metaphysics of invariance under intervention, by drawing upon the connections between Woodward, Russo and Nozick. – The aim was to highlight that the same logic of invariance under admissible transformations applies for Woodward's interventionist formulation of invariance. The main difference being that Woodward restrict the 'admissible transformations' to transformations that happen as a consequence of interventions (even though he does regard non-interventional changes in both the causal generalization and the background conditions).

Due to lack of space, there was no time to draw it all of the consequences nor go into all of the details, but the idea is that it is built into the logic of interventionism that the gradation of invariance leads to a gradation of metaphysical commitments, which in Woodward's terminology relates to the size of the domain of invariance serving as a metric for the amount of metaphysical commitments that he is committed to.

10 Literature list

- Baumgartner, M. (2012). The Logical Form of Interventionism. *Philosophia,40*(4), 751-761.
- Beebee, Helen, Christopher Hitchcock and Peter Menzies (2012). *The Oxford Handbook of Causation.* Oxford University Press
- Cartwright, N. (1979). Causal Laws and Effective Strategies. *Noûs, 13*(4), 419-437.
- Cartwright, N. (2003). Two Theorems on Invariance and Causality. *Philosophy of Science, 70*(1), 203-224.
- Collins, John, Ned Hall and L.A. Paul (2004). *Causation and Counterfactuals*. The MIT Press
- Danks, D. (2015). Goal-dependence in (scientific) ontology. *Synthese,192*(11), 3601-3616.
- Frisch, M. (2014). Physics and the Human Face of Causation. *Topoi, 33*(2), 407-419.
- Glynn, L. (2013). Causal foundationalism, physical causation, and difference-making. *Synthese, 190*(6), 1017-1037.
- Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*(5), 725-740.
- Godfrey-Smith, P., & Lesnikowski, Adam P. (2006). Theories and Models in Metaphysics. *The Harvard Review of Philosophy, 14*(1), 4-19.
- Hitchcock, C. (2012). Events and times: A case study in means-ends metaphysics. *Philosophical Studies, 160*(1), 79-96.
- Jantzen, B., Mayo, C., & Patton, D. (2015). Ontology & methodology. *Synthese, 192*(11), 3413-3423.
- Ney, A. (2009). Physical Causation and Difference-Making. *The British Journal for the Philosophy of Science, 60*(4), 737-764.
- Northcott, Robert. (2009). Is actual difference making actually different? *The Journal of Philosophy, 106*(11), 629-633.
- Nozick, Robert (2001). *Invariances – the Structure of the Objective World.* Harvard University Press
- Paul, L. (2012). Metaphysics as modeling: The handmaiden's tale. *Philosophical Studies, 160*(1), 1-29.
- Pearl, Judea (2000). *Causality – Models, reasoning and Inference.* Cambridge University Press.
- Russo, F. (2014). What Invariance Is and How to Test for It. *International Studies in the Philosophy of Science, 28*(2), 157-183.
- Schaffer, Jonathan. (2005). Contrastive Causation. *Philosophical Review,114*(3), 327.
- Strevens, M. (2007). Review of Woodward, Making Things Happen *. *Philosophy and Phenomenological Research, 74*(1), 233-249.
- Strevens, M. (2008). Comments on Woodward, Making Things Happen. *Philosophy and Phenomenological Research, 77*(1), 171-192.

- Waters, C. (2007). Causes That Make a Difference. *The Journal of Philosophy, 104*(11), 551-579.
- Woodward, J. (1997). Explanation, invariance, and intervention. *Philosophy of Science, 64*(4), 26-41.
- Woodward, J. (2000). Explanation and invariance in the special sciences. *The British Journal for the Philosophy of Science, 51*(2), 197-254.
- Woodward, J. (2007). Causation with a Human Face. In Price, H., & Corry, Richard (Eds.) *Causation, physics, and the constitution of reality : Russell's republic revisited*. (66-105). Oxford: Clarendon.
- Woodward, J. (2008). Response to Strevens*. *Philosophy and Phenomenological Research, 77*(1), 193-212.
- Woodward, J. (2013). Laws, causes, and invariance*. I Mumford, S., & Tugby, M. (Eds.) *Metaphysics and science* (Mind association occasional series). Oxford: Oxford University Press.
- Woodward, J. (2015). Methodology, ontology, and interventionism. *Synthese,192*(11), 3577-3599.
- Woodward, J. (2016). The problem of variable choice. *Synthese, 193*(4), 1047-1072.
- Woodward, J. (forthcoming). A Functional Account of Causation; or, A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard That Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment). *Philosophy of Science, 81*(5), 691-713.