THE ROUTLEDGE COMPANION TO THOUGHT EXPERIMENTS

Edited by
Michael T. Stuart, Yiftach Fehige and James Robert Brown

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published in 2018
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

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

Library of Congress Cataloging-in-Publication Data
A catalog record for this book has been requested

ISBN: 978-0-415-73508-7 (hbk)
ISBN: 978-1-315-17502-7 (ebk)

Typeset in Goudy
by Book Now Ltd, London

26
THOUGHT EXPERIMENTS AND IDEALIZATIONS

Julian Reiss

1 Introduction

Scientists rarely experiment on "nature" herself. Instead, they devise systems that are simplified, shielded from outside influences, arranged deliberately in ways that differ significantly from their naturally found analogues, made manageable and that stand in for natural systems before performing an experiment and making an inference to a principle that is hoped to extend beyond the experimental set up.

The deliberate distortions of nature that aim to address questions more effectively than asking nature directly would achieve are referred to as idealizations. They can occur in physical space, as when a glass tube is evacuated in order to prevent a cathode ray from ionizing gas molecules. But they can, and very frequently do, also occur in hypothetical space, when they are performed not on a chunk of nature but on a *representation* of it.

Scientific models idealize, often in order to make a problem mathematically tractable. This chapter considers idealizations in a context about which much less has been written, thought experiments. In what follows, I shall first introduce a small number of examples of idealizing thought experiments from physics, economics and biology. I will then distinguish a number of different kinds of idealizations one finds in these thought experiments and argue that they are all made for essentially the same purpose. The main discussion of this chapter occurs in Section 4, which asks how and under what conditions idealizations in thought experiments can be justified. Section 5 concludes by drawing some consequences of the discussion of the chapter's main argument for estimations of the chances of finding successful thought experiments in different scientific domains.

2 Thought experiments in physics, economics, and biology: some examples

Thought experiments are, like concrete experiments, manipulations on very special systems – except that they obtain in thought rather than in reality. Consider Galileo's

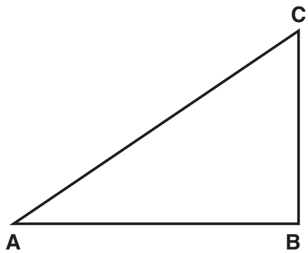

Figure 26.1

famous thought experiment of the inclined plane (Galilei 1967). In the thought experiment, Galileo constructs a system of two planes, one vertical (CB in Figure 26.1) and the other slanted (CA).

Along CA Galileo imagines a ball rolling, and along CB, a ball falling freely. As Salviati, Galileo's spokesman, says (Galilei 1967: 23),

> Now the line CA is meant to be an inclined plane, exquisitely polished and hard, upon which descends a perfectly round ball of some very hard substance. Suppose another ball, quite similar, to fall freely along the perpendicular CB.

The main point of the thought experiment is made when Salviati leads Sagredo, one of his interlocutors, into a contradiction by asking him about the speed of the balls. On the one hand, Sagredo agrees that both balls will have the same impetus (speed) when reaching points A and B, respectively – the speed required to take them back up to the same height. But on the other hand, it seems that the ball sliding down CA will have a lower speed than the falling ball. After all, CB is much shorter than CA, and the falling ball reaches B before the sliding ball reaches A. According to Thomas Kuhn, the chief lesson of this thought experiment is conceptual change (Kuhn 1981 [1963]). This is because the only way to resolve the tension of competing claims about the speeds of the balls is to distinguish different concepts of speed. There is, first, a concept of *instantaneous* velocity that measures a body's speed at a single point in time. And there is, second, a concept of speed that measures the time a body needs to traverse a given distance. As a body can be faster in one sense but not in the other, there is no contradiction.

Now consider a famous thought experiment from economics. In a seminal paper, George Akerlof seeks to explain the large price differential between new cars and cars that have just left the showroom (Akerlof 1970). Though Akerlof observes that there is a widely accepted explanation – people simply have a preference for new cars – he thinks that there is a better alternative. An important step in his argument that establishes his alternative explanation is the following thought experiment (489):

> Suppose (for the sake of clarity rather than reality) that there are just four kinds of cars. There are new cars and used cars. There are good cars and bad cars (which in America are known as "lemons"). A new car may be a good car or a lemon, and of course the same is true of used cars.

> The individuals in this market buy a new automobile without knowing whether the car they buy will be good or a lemon. But they do know that with probability q it is a good car and with probability (1 – q) it is a lemon; by assumption, *q* is the proportion of good cars produced and (1 – *q*) is the proportion of lemons.
>
> After owning a specific car, however, for a length of time, the car owner can form a good idea of the quality of this machine; i.e., the owner assigns a new probability to the event that his car is a lemon. This estimate is more accurate than the original estimate. An asymmetry in available information has developed: for the sellers now have more knowledge about the quality of a car than the buyers. But good cars and bad cars must still sell at the same price – since it is impossible for a buyer to tell the difference between a good car and a bad car. It is apparent that a used car cannot have the same valuation as a new car – if it did have the same valuation, it would clearly be advantageous to trade a lemon at the price of new car, and buy another new car, at a higher probability *q* of being good and a lower probability of being bad. Thus the owner of a good machine must be locked in. Not only is it true that he cannot receive the true value of his car, but he cannot even obtain the expected value of a new car.

It is the asymmetry in the information distribution that explains the price differential between new and used cars. This explanation is better than the original one because it does without an assumption about a specific preference individuals might or might not have. A structural explanation is preferable to one that requires an assumption about how individuals value certain goods. Moreover, as Akerlof indicates in an "applications" section of the paper, the new explanation can be reused in a variety of different contexts – insurance markets, labour markets, businesses where honesty matters, and credit markets in developing countries. His explanation is thus more unifying than positing a specific preference for new cars.

For a final example, let us turn to biology. Biology is rarely discussed in the literature on thought experiments. James Brown and Yiftach Fehige's otherwise comprehensive *Stanford Encyclopedia of Philosophy* entry, for example, leaves out biology completely (Brown and Fehige 2016). And yet, one can find thought experiments in evolutionary biology, population genetics, molecular biology, and artificial life and computational modelling in biology (Lennox 1991; Schlaepfer and Weber, this volume). Here is a famous one from Darwin's *Origin of the Species* (quoted from Lennox 1991, 228–9):

> Let us take the case of a wolf, which preys on various animals, securing some by craft, some by strength, and some by fleetness; and let us suppose that the fleetest prey, a deer for instance, had from any change in the country increased in numbers, or that other prey had decreased in numbers, during that season of the year when the wolf is hardest pressed for food. I can under such circumstances see no reason to doubt that the swiftest and slimmest wolves would have the best chance of surviving, and so be preserved or selected.
>
> Now, if any slight innate change of habit or of structure benefited an individual wolf, it would have the best chance of surviving and of leaving offspring. Some of its young would probably inherit the same habits or structure, and by the repetition of this process, a new variety might be formed which would either supplant or coexist with the parent-form of wolf.

According to James Lennox, thought experiments such as this do not play an evidential role; instead, they help to explore the explanatory potential of a theory, in this case the theory of natural selection.

3 Idealizations

Idealizations are ubiquitous in science and play a variety of roles. The models one finds in mathematical sciences such as theoretical physics, theoretical economics and theoretical biology often idealize in order to make problems mathematically tractable. For around 30 years, computer simulations have become more popular in these sciences, and computer simulations require different kinds of idealizations. They are, on the one hand, not constrained by the need to find closed-form solutions to a mathematical problem. But on the other, the more complex a problem becomes, the more computing time is needed, and computational resources are scarce. Thus different kinds of idealizations are introduced in order to save computing time. These kinds of "mathematical idealization" will not be considered further here, however, as thought experiments achieve their results without mathematization or computation.

Ernan McMullin considers what he calls "Galilean idealizations" and, among them, distinguishes "construct idealizations" and "causal idealizations." Both are "Galilean" in that they simplify the envisaged situation relative to the reality it stands in for. However, the former changes the conceptual representation of an object whereas the latter changes the problem situation itself (McMullin 1985, 255). An example of a construct idealization is Galileo's assumption that two weights suspended from a balance make right angles with the balance. The assumption is false since both weights are directed to the centre of the Earth. Among construct idealizations, McMullin further distinguishes between *formal* and *material* idealizations. Formal idealizations are false assumptions about features that are known to make a difference to the kind of explanation offered, but that are required in order to derive a result (258). Material idealizations concern features that are left unspecified in the envisaged situation because they are deemed irrelevant to the inquiry at hand (*ibid.*). This kind of idealization is also variously referred to as "Aristotelian" (Brown 2012; Frigg and Hartmann 2012) and "abstraction" (as opposed to idealization, see Cartwright 1989; Jones 2005).

Thought experiments are full of both formal and material construct idealizations. Galileo's thought experiment of the inclined plane, for instance, assumes "perfectly round" balls of "some very hard" substance (Galilei 1967, 23). There is no doubt that a less than perfectly round or soft ball would roll down the slope in a different manner (and so this is a formal idealization Galileo makes in order to derive his result). Given it is perfectly round and hard, it does not matter which colour the ball is or what it is made of. So Galileo specifies only that the substance is very hard and idealizes away all other properties it might have.

In Akerlof's thought experiment, there are only two categories of cars, "good" and "bad." Clearly, cars have more relevant properties than quality, quality is itself a vector rather than a scalar, and each dimension has more than just two values. Unlike in Galileo's physical inquiry, in an economic inquiry the colour of the object might matter a great deal. On the other hand, there are certainly some properties of cars that are not relevant to their tradability – the number of molecules they are made of, or whether they were made by a German or a Korean robot (assuming that that does not affect their quality!).

McMullin describes the reason for making causal idealizations as follows (264):

> The unordered world of Nature is a tangle of causal lines; there is no hope of a "firm science" unless one can somehow simplify the tangle by eliminating, or otherwise neutralizing, the causal lines which impede, or complicate, the action of the factors one is trying to sort out.

Causal idealization too comes in two forms: experimental and subjunctive. Experimental idealization is the contrivance of a simplified physical (or biological, social, etc.) system, designed to answer specific scientific questions or problems (265ff.). Subjunctive idealization contrives such a system in the imagination: it answers "what would happen if … ?" questions (268ff.).

All thought experiments involve causal idealizations of the subjunctive kind. Galileo assumes away air resistance and, implicitly, all other forces that might influence the speed of the balls. Akerlof, again, implicitly, keeps constant all factors that might influence individuals' propensities to buy and sell cars such as incomes, preferences, expectations and so on. Darwin's perhaps appears as the least idealized thought experiment, and it does not make any explicitly idealizing assumptions. However, it too has to assume that the relative increase of fleeting prey is the only cause of the change in the wolf's genetic pool. After all, if the reason for the decimation of the non-fleeting prey was, say, a virus that also befell swift and slim (but not other) wolves, we would not expect the formation of a new variety of wolf, or at least not of the same new variety Darwin imagines.

Among the subjunctive idealizations, McMullin further distinguishes between merely hypothetical and counterfactual ones, depending on whether or not the hypothesized idealized system could, in principle, be physically (biologically, socially, etc.) realized. It is certainly conceivable that a wolf territory can be controlled in such a way that an increase in fleeting prey is the only possible cause of changes in swift and slim wolves' fitness. On the other hand, perfect vacuums, spheres and perfectly smooth planes can, at best, be approximated.

4 Justifying idealizations

4.1 Weisberg on idealizations

Michael Weisberg argues that there are three different kinds of idealization in scientific practice, each coming with its own goals, rationale and justification (Weisberg 2007). The Galilean type, called "minimalist" by Weisberg, is only one among others. The other two I will refer to as Pythagorean and perspectivist, respectively.[1] Pythagorean assumptions aim to make a problem mathematically or computationally tractable. They are, according to Weisberg, justified pragmatically: "We simplify to more computationally tractable theories in order to get traction on the problem. If the theorist had not idealized, she would have been in a worse situation, stuck with an intractable theory" (641). This type is sensitive to developments in computational power and mathematical techniques; advances in the latter should bring about a degree of de-idealization. At any rate, different kinds of mathematics and computation techniques require different kinds of idealizing assumptions. In the ideal limit, Pythagorean idealizations are unnecessary. If the language of mathematics becomes infinitely flexible and computers infinitely powerful, there is no need to make false assumptions for the sake of mathematical or computational tractability.

Perspectival idealizations aim to represent an aspect of a phenomenon accurately. One and the same phenomenon can be regarded from different angles and different representations can bring out different aspects of a phenomenon. Different representations focus on different causal processes and provide different accounts of phenomena. No one explanation is the uniquely correct one. One reason for this is that researchers pursue different goals with their representations: explanatory power, accuracy, precision and so on. There are inherent trade-offs between these goals and so no representation can maximize all of them at the same time. Different idealizations help to achieve different goals (646).

In my view, there are no Pythagorean or perspectival idealizations in thought experiments. As intuitions and not mathematical deduction or computer-aided calculation drive the derivation of the result, no Pythagorean idealizations are needed, and I am not aware of any set of competing thought experiments that examine the same phenomenon.[2]

4.2 Analysis and synthesis

Typical idealizations in thought experiments are thus all of the "Galilean" type considered in Section 3. They all serve one and the same purpose: to apply the first (analytic) stage of the method of analysis and synthesis. A complex problem such as that of calculating the speed of a falling body in, as Carl Menger would call it, "full empirical reality" (Menger 1963 [1887]) is addressed by first translating it into a related but simpler problem that considers only one or a small number of main causes of the phenomenon of interest; then a principle is established that allows the explanation and prediction of the simplified phenomenon; a prediction or explanation of the original phenomenon is finally reached by adding in corrections for the omitted causes. The analytic stage of the method proceeds by separating individual causal lines from each other and establishing a principle for each one. The synthetic stage combines the different principles to make a prediction or explanation of a complex phenomenon.

The fruitfulness of the method of analysis and synthesis, and of using thought experiments to implement its first stage, depends on whether or not two conditions are met: an epistemic and an empirical condition. The epistemic condition is that there are reliable methods for finding out what a single causal line does in isolation. In the present context, the question is whether thought experimentation can be justified as such a method. The empirical condition is that single causal lines make a systematic contribution to the outcome of interest that persists when "accidents" or disturbing causes are present. This is independent of whatever method is used, but information about whether the second condition is met or not will help to explain where certain methods work, why they fail and why we find a preponderance (or dearth) of certain methods in certain domains.

4.3 Thought experiments and the epistemology of idealizations

The success of experiments often turns on researchers' ability to make effective experimental idealizations. Peter Achinstein discusses Heinrich Hertz's attempts to determine whether cathode rays are electrically charged (Achinstein 2001). In one experiment, he introduced oppositely electrified plates into the cathode tube and reasoned that if cathode rays were electrically charged they should be deflected by these plates. The deflection did not occur and Hertz inferred that cathode rays are electrically neutral. The reason for the absence of a visible deflection was not, however, that the rays were not charged, as J. J. Thomson found out fourteen years later. He reasoned that if they were charged, they would ionize the gas molecules in the cathode tube, producing positive and negative charges that will neutralize the charge on the metal plates between which the cathode rays travel. Indeed, repeating Hertz's experiment with a properly evacuated tube, Thomson could observe the deflection of the rays. Thus, the failure of an experimental idealization can be responsible for the misinterpretation of the result of the experiment. If the idealizations work, however, there is not much uncertainty about the results as it is produced by nature.

A thought experiment has opposite characteristics. There can be little doubt about the success of a subjunctive idealization – we cannot fail to assume away air resistance or to make a plane perfectly smooth. We might fail to think of every possible cause of an outcome of interest, but even in this case the catchall assumption, "No other causes affect the outcome," will do the trick. On the other hand, since it is reached by intuition, there can be considerable uncertainty about the accuracy of the result. Who is to guarantee that nature, if asked the same question in a well-designed and executed experiment, would give the same answer? And how do we know how nature would answer in the first place?

Richard Laymon argues that idealizing assumptions in a thought experiment can be defended in either of two ways (Laymon 1991): there exists a series of (real/material) experiments such that the idealized situation is asymptotically approachable or there exist experiments and good theories of the disturbing causes so that their influence on the result can be calculated. Theories of disturbing forces are not normally available in the contexts where thought experiments play important roles. An early theory of air resistance is due to Newton who published it about a century after Galileo's *Dialogue*. In economics there are no quantitative theories of disturbing causes that I can think of. Series of experiments that asymptotically approach idealized set-ups, on the other hand, are plentiful. Half a century before Galileo published his *Dialogue*, Simon Stevin, the Flemish physicist and mathematician, refuted Aristotle's theory of falling bodies empirically (quoted from Kühne 2005, 32):

> The experience against Aristotle is the following: Let us take (as the very learned Mr. Jan Cornets de Groot, most industrious investigator of the secrets of Nature, and myself have done) two spheres of lead, the one ten times larger and heavier than the other, and drop them together from a height of 30 feet on to a board or something on which they give a perceptible sound [*merckelick gheluyt*]. Then it will be found that the lighter will not be ten times longer on its way than the heavier, but that they fall together on to the board so simultaneously that their two sounds seem to be one and the same rap. The same is found also to happen in practice with two equally large bodies whose gravities are in the ratio of one to ten; therefore Aristotle's aforesaid proportion is incorrect.

Galileo's *Dialogue* contains similar experimental results (in addition to his famous thought experiments). It is also at least conceivable that he experimented with slopes of an increasing degree of smoothness. Akerlof's argument is based on the idea that people do not leave massive arbitrage opportunities unexploited. This again is something for which a good deal of empirical evidence existed at the time Akerlof published his thought experiment.

Margaret Schabas has an interesting theory of thought experiments in economics, a discussion of which will bring out another aspect of the epistemology of idealizations (Schabas 2008, this volume). According to her, bona fide thought experiments usually have two characteristics that set them apart from related techniques such as mathematical demonstration, modelling or narration. First, they involve a genuinely experimental moment such as a manipulation or intervention. Second, they begin with a jarring counterfactual "that transports the mind to a different and distant world, as opposed to a proximate alternative world" (Schabas 2008, 162). One of Hume's famous monetary thought experiments is a case in point (quoted from Schabas 2008, 161):

> For suppose, that, by miracle, every man in GREAT BRITAIN should have five pounds slipt into his pocket in one night; this would much more than double the whole money that is at present in the kingdom; yet there would not next day, nor for some time, be any more lenders, nor any variation in the interest [rate].

Schabas argues that Hume used this thought experiment to make the "in principle" neutrality of money and its "in practice" non-neutrality compatible with each other. Milton Friedman's present-day analogy achieves something similar (Friedman 1969, 4–5):

> Let us suppose now that one day a helicopter flies over this community and drops an additional $1,000 in bills from the sky, which is, of course, hastily collected by members of the community. Let us suppose further that everyone is convinced that this is a unique event which will never be repeated.

Both are clearly thought experiments according to Schabas's criteria as there is a quasi-experimental manipulation and the initial counterfactual is one which is jarring: neither overnight miraculous doublings of the money stock nor helicopter drops of large quantities of money occur in real economies, not even approximately. In my view, these thought experiments aim to achieve the same as those we have considered so far by making one dramatic idealization instead of a number of less conspicuous ones: they isolate a single causal line. In Hume's and Friedman's thought experiments, this happens by artificially increasing the size of the causal factor of interest; whatever other factors then might contribute to the outcome, it will be negligible relative to the main factor. To use Marcel Boumans's apt terminology, whereas the thought experiments from Section 2 require "*ceteris absentibus*" assumptions, due to the causal effect size of the idealization in Hume's and Friedman's thought experiments, only a "*ceteris neglectis*" assumption concerning the other factors has to be made (Boumans 2005, 118–19; see also Musgrave 1981 on assumptions in economic models and Morgan 2013 on "swamping" causes that help the analysis of natural experiments).

Historians often use counterfactuals in order to establish singular causal explanations. In order to establish, say, whether the assassination of Archduke Franz Ferdinand played a causal role in the outbreak of World War I or the Greeks' victory in the battle of Salamis in the development of the hegemony of Western values, they ask if the war had broken out if Franz Ferdinand had not been assassinated or if Western values had still risen to dominance if the Greeks had lost (Hanson 2006; Lebow 2014). In order to assess these counterfactuals, historians use a "historical consistency" criterion (Tetlock and Belkin 1996; Reiss 2009, 2012). Not too much history is supposed to be rewritten in order to make the counterfactual antecedent true. Importantly, historians ask whether the counterfactual antecedent is "co-tenable" with what else is known about history: the counterfactual antecedent *A* is co-tenable with a proposition *S* if it is not the case that *S* would be false if *A* were true (Goodman 1954, 15). So to assess whether an antecedent is historically consistent, they test whether making the antecedent true would violate firmly held beliefs about historical events and patterns. Hume's counterfactual is, arguably, not historically consistent in this sense. For miracles to occur, many beliefs about history would have to be different.

The rationale behind this aside on historiographic methodology was to draw attention to one of the historians' reasons for not rewriting history too much when making counterfactual thought experiments, which is an epistemic reason (Lebow 2010, 55, original emphasis):[3]

> Max Weber insisted that plausible counterfactuals should make as [minimal] historical changes as possible on the grounds that the more we disturb the values, goals, and contexts in which actors operate, the less predictable their behavior becomes. Counterfactual arguments make a credible case for a dramatically different future on the basis of one small change in reality are very powerful and the plausible rewrite rule should be followed whenever possible. The *nature* of the changes made by the experiment are nevertheless more important than the *number* of changes. A plausible rewrite that makes only one alteration in reality may not qualify as a plausible world counterfactual if the counterfactual is unrealistic or if numerous subsequent counterfactual steps are necessary to reach the hypothesized consequent. A counterfactual based on several small changes, all of them appearing plausible, may be more plausible, especially if they lead more directly to the consequent.

We do not know how people would react if their wealth suddenly doubles overnight. Hence, according to historians in the Weber tradition, we should not ask what would happen if it were to double. The historians' criterion is not exactly the same as Laymon's. A historically consistent counterfactual might be one for which there is no series of asymptotically approaching experiments. And even if we have such a series of experiments, the counterfactual might still be historically inconsistent because it is an alteration of the course of history that is, even though generally possible, implausible in a given historical context.[4] Both criteria can help us to determine whether a given idealization meets the epistemic condition.

4.4 Idealization and causality

The second condition for the method of analysis and synthesis to be used fruitfully is empirical: causes have to combine additively, at least approximately. If, say, the linear air resistance hypothesis is true, according to which the force due to air resistance is proportional to speed and acts in the direction opposite to motion, the total force acting on a falling body is given by:

$$F_{tot} = F_{grav} + F_{air} = mg - ks(t),$$

where F_{tot} is the total force, F_{grav} the force due to the Earth's gravitational field, F_{air} air resistance, m the body's mass, g the gravitational acceleration, k a constant and $s(t)$ the body's speed at t. In this case, both forces make simple, additive contributions to the total force. Ronald Laymon describes the advantages of this type of behaviour as follows (Laymon 1995, 361):

> In summary, additivity has two significant virtues. First, it allows ignored causes to be treated as if they had value zero. Second, it allows for the easy transfer of causal information from one situation to another because the effects of a variation in cause remain the same regardless of what combination of other causes is present.

To see this, suppose, by contrast, that the force due to gravity interacts with air resistance. What I mean by "interaction" is that the functional form of the relation between speed and the Earth's gravity and thus the size of the contribution gravity makes to the total force, is itself dependent on the value of force due to air resistance. The amount of air resistance an object experiences depends on its speed, its cross-sectional area, its shape and the density of the air. In the extreme, we could have not one law of falling bodies but many: one law for slim and swift bodies, one for large and laggard, one for the English wet summer (as humidity and temperature affect the density of the air; and of course there would be different English wet summer laws for swift/slim bodies and for large/laggard ones), one for the dry Californian winter and so on and so forth. This is clearly not how falling bodies behave. That they do not behave this way is something we learn from empirical evidence, and Galileo was well aware of the empirical facts before presenting his thought experiment.

Not all things are as regularly behaved as falling bodies. To give one example from social science, the causes of economic growth are manifold. There was a time when many policy makers and economists subscribed to the "Washington consensus" – a set of economic policy prescriptions made to help crisis-ridden developing countries trigger and sustain growth. The original list was the following (Rodrik 2007, 17):

- A Fiscal discipline
- B Reorientation of public expenditures
- C Tax reform
- D Interest rate liberalization
- E Unified and competitive exchange rates
- F Trade liberalization
- G Openness to foreign direct investment
- H Privatization
- I Deregulation
- J Secure property rights

It would be a mistake to assume that each of these recommendations corresponds to a factor that has an independent additive effect on growth. In fact, looking at recent empirical evidence, it looks as though there is no correlation or a negative correlation between a developing country's success at implementing these policies and its economic success – "best practice" countries (such as some Latin American countries) have done fairly poorly, and countries that have a poor record in at least some of these dimensions have done very well (such as China).

After a look at the data from a large set of developing countries, Dani Rodrik assesses the situation as follows (Rodrik 2007, 39): "No country has experienced rapid growth without minimal adherence to what I have termed higher-order principles of sound economic governance – property rights, market-oriented incentives, sound money, fiscal solvency. But … , these principles have often been implemented via policy arrangements that are quite unconventional." Thus, some of the items of the Washington consensus seem to be necessary conditions for growth. However, what matters is that a country gets the specific mix right, and what that means is very different from country to country. If Rodrik is correct about this, the causes of economic growth are such that the contribution a factor makes to the outcome depends on the background constellation of other factors within which it operates.

4.5 Causality, metaphysics and semantics

Is what I called the "empirical condition" in fact a metaphysical condition? Does it tell us anything about the nature of causality? It is certainly true that the most prominent attempts to make sense of this aspect of scientific methodology is often underwritten with a form of Aristotelian realism about causal powers or capacities (see in particular Cartwright 1999, especially ch. 4, and Mumford and Anjum 2011). Aristotelian realism is particularly popular among social scientists (see, for instance, Pawson 2006 and Groff 2008, part III). But as I have argued in previous work, the idea that complex situations can be analyzed by examining the laws that describe what individual factors do in isolation and then to predict what happens on the basis of these laws plus a law of composition is in fact metaphysically neutral. For one thing, apart from Aristotelian realist, there are also Kantian (Watkins 2005) and fictionalist (Vaihinger 1924) interpretations of causal contributions available. For another, what matters is *whether* the method of analysis and synthesis works, not *why* it works when it does. That factors have powers or capacities that make stable causal contributions is at best a sufficient condition for the method to work, but it is not necessary. For the method to work it would be entirely sufficient if the factors behaved *as though* they had causal powers/capacities.

In my view, the question is more one of semantics. Whether or not a causal factor makes stable contributions (or behaves as though it does) makes an enormous difference to scientists' inferential practices and abilities. On the one hand, it makes a difference to the methods with which causal claims are established – in particular to the kinds of idealizations these methods can properly make. Galilean idealizations work best when factors (behave as though they) make stable causal contributions. On the other hand, it makes a difference to the inferences that the acceptance of the causal claim entitles scientists to make. A claim that ascribes to a factor the power/capacity to make stable causal contributions allows inferences about situations that differ significantly from the (thought) experimental situation, whereas a contextualist causal claim does not license such inferences. To the extent that the meaning of causal claims is related to these inferential practices and abilities, the applicability of the Galilean method has therefore conceptual consequences (for a defence of such an inferentialist interpretation of causal claims, see Reiss 2015).

It certainly did so for John Stuart Mill. Mill would only call factors causes if they had the property of additivity (Laymon 1995, 360):

> Mill wanted more: that something not be a cause unless it had this property. In a nutshell, Mill's idea was that being a cause requires an independence of potency that can only exist if the efficacy of the cause does not depend on the values of other descriptive or causal parameters.

Laymon endorses Mill's view in the end (371), but it seems overly strong. Causes are difference makers. Their presence should make a difference to the outcome in the context in which the outcome obtains. Whether or not it continues to be a difference maker in other contexts, and in particular, whether the size of the difference it makes is stable across contexts should not affect our conceptualization of a factor as a cause (though it can of course affect our conceptualization of a factor as a particular kind of cause). All difference-making accounts of cause such as regularity accounts (e.g., Mackie 1980), counterfactual accounts (e.g., Lewis 1973) and interventionist accounts (e.g., Woodward 2003) agree with this reading.

5 Conclusions: thought experiments and their proper domain

Even though above I have described an example from physics where causes are additive and one from economics where they are not, it is not the case that one can generally assume that physics causes are additive and causes in the special sciences are not. Forces are additive, but many other causes in physics are not (for a detailed discussion of additivity in physics and its relation to idealization, see Laymon 1995). There is evidence that at least some economics causes have this property (Reiss forthcoming).

Nevertheless it is true that the method of analysis and synthesis appears to work much better in some domains than in others. In the natural sciences, it has been popular and successful since its Galilean origins. In the social and behavioural sciences, it is popular and being used, but often not very successfully. As McMullin remarks (McMullin 1985, 267–8):

> We have become accustomed in recent decades to the charge that this sort of ... idealization can, in the context of the behavioral sciences, distort the behavior one is trying to understand. Ethologists have urged that the "Galilean methods" advocated by behaviorists and others in biology and psychology inevitably alter the behavior to be studied in ways that undermine the value of these methods as guides to genuine scientific truth. The technique of isolating causal lines does not seem to work so well when the object under study is a complex organism or a social group. The "impediments" cannot be defined and eliminated as they can in the case of a ball on an inclined plane.

Perhaps this explains why we find thought experiments more frequently in some sciences rather than others. All thought experiments make Galilean idealizations, and these will be successful only to the extent that the two conditions discussed in Section 4 are met. As McMullin argues, the second condition is unlikely to be met in the social and behavioural sciences, and there are reasons to believe that the same is true for the epistemic condition. Schabas suggests that thought experiments are rare in economics. If what I have been argued in this chapter is correct, she may well be right.[5]

Notes

1 Weisberg calls them "Galilean" and "multiple model idealizations." The former I find rather misleading, as Galileo himself clearly aimed at providing accounts of single causal lines, which, according to Weisberg, is the main purpose of what he calls "minimalist" idealizations. That some of the principles that describe the behaviour of single causal lines can be described in mathematical terms is true, but the reason for the idealization is to get to the single causes. Mathematical description is a by-product. This is entirely different, for instance, in contemporary mathematical economics where idealizations are frequently introduced in order to make a problem mathematically tractable whether or not the outcome is a representation of a single cause. If a production function is assumed to be Cobb-Douglas, this assumption is made for mathematical convenience, not in order to focus on an isolated cause and at the expense of what Galileo called accidents. Because of their goal to make a problem mathematically (or computationally) tractable, I call these non-Galilean idealizations "Pythagorean."

2 John Norton examines what he calls "thought experiment – anti-thought experiment pairs" (Norton 2004). But his purpose is to show that it is possible to find convincing pairs of narratives so that one demonstrates the truth of a claim and the other its falsehood. The different thought experiments thus have the *same* goal but *opposite* results rather than different goals.

3 I substituted a "minimal" for the original "few" because it is more consistent with what Lebow says later on in the quote. I also omitted a footnote that provides a reference to Max Weber's (1905).

4 An example for such a historically inconsistent albeit possible counterfactual might be, "Had Chamberlain confronted Germany about the Sudetenland ..." There are certainly situations in history in which an aggressor state was confronted by a large imperial power. And yet, given what we know about Chamberlain, it is most unlikely that *he* would ever have done so. See Khong (1996) for a discussion.

5 Having a more liberal understanding of the term I do not think that thought experiments are quite as rare as Schabas suggests. However, I entirely agree that *successful* thought experiments in economics are rare, for the reasons given.

References

Achinstein, P. (2001) *The Book of Evidence*, Oxford: Oxford University Press.

Akerlof, G. (1970) "The market for 'lemons': Quality uncertainty and the market mechanism," *Quarterly Journal of Economics* 84: 488–500.

Boumans, M. (2005) *How Economists Model the World Into Numbers*, London: Routledge.

Brown, J. R. (2012) "What do we see in a thought experiment?" in *Thought Experiments in Science, Philosophy, and the Arts*, edited by M. Frappier, L. Meynell, and J. R. Brown, London: Routledge.

Brown, J. R. and Fehige, Y. (2016) "Thought experiments," *Stanford Encyclopedia of Philosophy*. Available here: http://plato.stanford.edu/archives/spr2016/entries/thought-experiment/ (accessed 15 November 2016).

Cartwright, N. (1989) *Nature's Capacities and Their Measurement*, Oxford: Clarendon.

Cartwright, N. (1999) *The Dappled World*, Cambridge: Cambridge University Press.

Friedman, M. (1969) *The Optimum Quantity of Money and Other Essays*, Chicago: Aldine Publishers.

Frigg, R. and Hartmann, S. (2012) "Models in science," *The Stanford Encyclopedia of Philosophy*. Available here: http://plato.stanford.edu/entries/models-science/ (accessed 15 November 2016).
Galilei, G. (1967) *Dialogue Concerning the Two Chief World Systems – Ptolemaic and Copernican*, translated by S. Drake, Berkeley: University of California Press.
Goodman, N. (1954) *Fact, Fiction and Forecast*, Atlantic Highlands: Athlone Press.
Groff, R. (2008) *Revitalizing Causality: Realism about Causality in Philosophy and Social Science*, London: Routledge.
Hanson, V. D. (2006) "A stillborn west? Themistocles at Salamis, 480 BC," in *Unmaking the West: "What-if?" Scenarios that Rewrite World History*, edited by P. Tetlock, R. N. Lebow, and G. Parker, Ann Arbor: University of Michigan Press.
Jones, M. (2005) "Idealization and abstraction: A framework," in *Idealization XII: Correcting the Model – Idealization and Abstraction in the Sciences*, edited by M. Jones and N. Cartwright, Amsterdam: Rodopi.
Khong, Y. F. (1996) "Confronting Hitler and its consequences," in *Counterfactual Thought Experiments in World Politics*, edited by P. Tetlock and A. Belkin, Princeton: Princeton University Press.
Kuhn, T. (1981 [1963]) "A function for thought experiments," in *The Structure of Scientific Revolutions*, Oxford: Oxford University Press.
Kühne, U. (2005) *Die Methode des Gedankenexperiments*, Frankfurt: Suhrkamp.
Laymon, R. (1991) "Thought experiments by Stevin, Mach and Gouy: Thought experiments as ideal limits and as semantic domains," in *Thought Experiments in Science and Philosophy*, edited by T. Horowitz and G. Massey, Savage: Rowman & Littlefield.
Laymon, R. (1995) "Experimentation and the legitimacy of idealization," *Philosophical Studies* 77: 353–375.
Lebow, R. N. (2010) *Forbidden Fruit: Counterfactuals and International Relations*, Princeton: Princeton University Press.
Lebow, R. N. (2014) *Archduke Franz Ferdinand Lives!: A World Without World War I*, New York: Palgrave Macmillan.
Lennox, J. (1991) "Darwinian thought experiments: A function for just-so stories," in *Thought Experiments in Science and Philosophy*, edited by T. Horowitz and G. Massey, Savage: Rowman & Littlefield.
Lewis, D. (1973) "Causation," *Journal of Philosophy* 70: 556–567.
Mackie, J. (1980) *The Cement of the Universe: A Study of Causation*, 2nd ed., Oxford: Oxford University Press.
McMullin, E. (1985) "Galilean idealization," *Studies in the History and Philosophy of Science* 16: 247–273.
Menger, C. (1963 [1887]) *Problems of Economics and Sociology*, translated by F. J. Nock, Urbana: University of Illinois Press.
Morgan, M. (2013) "Nature's experiments and natural experiments in the social sciences," *Philosophy of the Social Sciences* 43: 341–357.
Mumford, S. and Anjum, R. L. (2011) *Getting Causes from Powers*, Oxford: Oxford University Press.
Musgrave, A. (1981) "'Unreal assumptions' in economic theory: The F-twist untwisted," *Kyklos* 34: 377–387.
Norton, J. (2004) "Why thought experiments do not transcend empiricism," in *Contemporary Debates in the Philosophy of Science*, edited by C. Hitchcock, Oxford: Blackwell.
Pawson, R. (2006) *Evidence-Based Policy: A Realist Perspective*, London: Sage.
Reiss, J. (2009) "Counterfactuals, thought experiments and singular causal analysis in history," *Philosophy of Science* 76: 712–723.
Reiss, J. (2012) "Counterfactuals," in *The Oxford Handbook of the Philosophy of Social Science*, edited by H. Kincaid, Oxford: Oxford University Press.
Reiss, J. (2015) *Causation, Evidence, and Inference*, New York: Routledge.
Reiss, J. (forthcoming) "Are there social scientific laws?," in *The Routledge Companion to Philosophy of Social Science*, edited by L. McIntyre and A. Rosenberg, London: Routledge.
Rodrik, D. (2007) *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*, Princeton: Princeton University Press.

Schabas, M. (2008) "Hume's monetary thought experiments," *Studies in History and Philosophy of Science Part A* 29: 161–169.
Tetlock, P. and Belkin, A. (eds) (1996) *Counterfactual Thought Experiments in World Politics: Logical, Methodological and Psychological Perspectives*, Princeton: Princeton University Press.
Vaihinger, H. (1924) *The Philosophy of 'As If': A System of the Theoretical, Practical and Religious Fictions of Mankind*, London: Routledge and Keagan Paul.
Watkins, E. (2005) *Kant and the Metaphysics of Causality*, Cambridge: Cambridge University Press.
Weber, M. (1905 [1949]) "Objective possibility and adequate causation in historical explanation," in *The Methodology of the Social Sciences*, edited by M. Weber, E. Shils, and H. Finch, Glencoe: Free Press.
Weisberg, M. (2007) "Three kinds of idealization," *Journal of Philosophy* 104: 639–659.
Woodward, J. (2003) *Making Things Happen*, Oxford: Oxford University Press.