AARHUS UNIVERSITY

Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Bergenholtz, C., & Busch, J. (2015). Self-Fulfillment of Social Science Theories: Cooling the Fire. *Philosophy of the Social Sciences*, 46(1), 24-43. doi:10.1177/0048393115619233

Publication metadata

Title:	Self-Fulfillment of Social Science Theories: Cooling the Fire
Author(s):	Bergenholtz, C., & Busch, J.
Journal:	<i>Philosophy of the Social Sciences</i> , 46(1), 24-43.
DOI/Link:	http://journals.sagepub.com/doi/abs/10.1177/0048393115619233
Document version:	Accepted manuscript (post-print)

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

This is a post-print version of the paper

Self-fulfillment of social science theories: Cooling the fire

by Carsten Bergenholtz¹ & Jacob Busch²

¹ Carsten Bergenholtz, Department of Management, Aarhus BSS, Aarhus University, Bartholins Alle 12, 8000 Aarhus, Denmark

² Jacob Busch, Department of Culture and Society, Aarhus University, Jens Chr. Skous Vej 7, 8000 Aarhus, Denmark

Corresponding author:

Carsten Bergenholtz, Department of Management, Aarhus BSS, Aarhus University, Bartholins Alle 12, 8000 Aarhus, Denmark

Email: cabe@asb.dk

Abstract

Self-fulfillment of theories is argued to be a threat to the social science in at least two ways. Firstly, a realist might worry that self-fulfillment constitutes a threat to the idea that social science is a proper science consistent with a realist approach that develops true and successful statements about the world. Secondly, one might argue that the potential self-fulfilling nature of social science theories potentially undermines the ethical integrity of social scientists. We argue that if one accepts that social science theories are not based on laws akin to those that govern natural reality *or* acknowledges that if one can predict self-fulfillment via a meta-theory that explains the underlying regularities of the self-fulfilling change, the threat to realism is dismantled. Furthermore, on the basis of these arguments we show that if one is not able to predict the (moral) consequences of a theory, it is difficult to ascribe moral responsibility at the individual level. It is therefore not the potential self-fulfillment of theories per se that poses an ethical challenge, in contrast to claims in the literature.

Keywords

Self-fulfillment, performativity, scientific realism, objectivity, science ethics

1. Introduction

Some of the primary aims of social science are that it should be able to explain, predict, inform and influence social behavior. In an extension of this view there has recently been a resurgent and heated interest as to whether social science theories not merely inform and influence social behavior but are to be characterized as self-fulfilling prophecies that might turn false to true (Callon 1998; Ferraro et al. 2005; MacKenzie 2006; Ferraro et al. 2009; Felin & Foss 2009a; Felin and Foss 2009b; Santos and Rodrigues 2009; Vosselman 2014; Mason et al. 2015). The most well-known illustration of this issue is Merton's (1948) bank-run example: At a particular time t_1 , a bank is not insolvent, while at a later time t_2 , the bank is prophesized to be so and this prophecy is made publicly available. At a later time t_3 the bank has now become insolvent, since customers are withdrawing all their money, which given the information at t_2 , can be considered rational behavior. The prophecy might in this way turn false to true in the short term (see Felin & Foss 2009b: p. 5 for a discussion of the temporal dynamics). The example however makes it clear, that the sense in which the prophecy 'turns false to true' not only depends on aspects inherent to it qua theory, but also depends on institutional features and e.g. the legitimacy and social influence of the one proposing the statements (Sherif 1937). Other known examples are the Black-Scholes-Merton formula which influenced financial markets (see e.g. Callon 1998 and MacKenzie 2006)¹, how learning the economic notion of self-interest might make economic students more selfish (Cipriani, Lubian, and Zago 2009) or how the performance of (e.g. black) students might depend on racial expectations (Ferraro, Pfeffer, and Sutton 2009; Aronson, Fried, and Good 2002).

The overall terminology and structure of this debate is in part derived from classical debate between (social) constructionism at one end of the spectrum and realism on the other (Boghossian 2006). On the one hand, social constructionists argue that 'theories shape the world' (Latour and Woolgar 1979) and show how institutions, norms and language play a significant role in this process (Ferraro et al. 2005). On

¹ Although, the Black-Scholes-Merton formula mainly formalized and branded an existing practice. They were thus more entrepreneurial than theory developers (Haug and Taleb 2011).

the other hand, realists argue that ‘reality is out there’ (Devitt 1984) and that real theoretical mechanisms and regularities can not be changed by putting false theories forward (Felin and Foss 2009a).

Depending on the aforementioned philosophy of science perspective, self-fulfillment of theories is argued to be a threat to social science in at least two ways. Firstly, a realist might worry that self-fulfillment is questioning whether social science conflicts with scientific realism²; in brief the view that scientific theories offer true knowledge of objective reality, and that we have reason for believing that they are true. This position would argue that for a science to count as being properly scientific, it has to allow for a scientific realist interpretation, and so by implication, a violation of scientific realism might constitute a threat to the idea that social science is a proper science that develops true and successful statements about the world (eg. Felin and Foss (2009ba)). The threat that self-fulfillment may pose to scientific realism seems to be a threat that is unique to the social sciences in contrast to the natural sciences, where scientific theories are believed to be aimed at tracking reality and to have no impact upon their object of study, thereby rendering self-fulfillment of theories impossible (Taylor 1983; MacKenzie 2006; Callon 1998; Bourdieu 1992; Merton 1948). Secondly, one might argue that the potential self-fulfilling nature of social science theories creates a certain moral responsibility for the individual researcher, since even wrong theories might be self-fulfilled in the future (Ferraro et al. 2005; Ferraro et al. 2009). Thus researchers involved in theory development are potentially responsible for changing the world in accord with how they theorize that the world is, and are in that respect engaged in a world changing activity for which they can be held ethically accountable. These two threats constitute the motivation for the present paper.

We intend to clarify the nature of the two threats. More specifically, we will argue that if one accepts that social science theories are not based on a conception of laws derived from the natural sciences *or* acknowledges that if one can predict self-fulfillment via a meta-theory that explains the underlying regularities of the self-fulfilling change, the threat to scientific realism can be dismantled. Furthermore,

² Or just realism, in short.

we will argue that self-fulfillment does not pose a particular ethical challenge to the individual researcher but should be addressed at the level of the scientific community. The overall aim with this short paper is to cool the fire in the debate between realists and social constructionists in order to dismantle the challenges of self-fulfillment and discuss how self-fulfillment constitutes an ethical challenge on a scientific community level.

First we clarify the territory in order to set the battleground. Hereafter we extend the present conceptions of scientific realism in the self-fulfillment literature by presenting different dimensions of scientific realism, in order to frame the topic properly in the philosophy of science. This involves a clarification of the nature of self-fulfillment, and the introduction of time-operators and meta-level theories. Finally, the potential ethical implications of self-fulfillment are discussed.

2. Clarifying the territory

2.1. Terminology

A vital and confusing feature of the debate over self-fulfillment of theories is the variety of terms and perspectives that to some extent are applied interchangeably. This confusion exists on two different levels.

First of all, one can talk about *the nature of change* that a theory brings about i.e. either change due to performativity, self-fulfillment or even self-falsification. Various forms of performativity have been introduced to the debate, with generic performativity merely implying that economic phenomena are not mirroring Platonic ideals or natural categories, but performed by human, economic practices (MacKenzie 2004). The adoption of a theory or economic model impacts economic practice, but does not necessarily improve the fit between theory and economic practice – just as insight into chemical reactions might improve the ability to handle chemicals, but will not improve the fit between the theory on chemical reactions and real world phenomena.

Stronger versions of performativity (Austinian or Barnesian) argue that the adoption of an economic theory transforms economic practice, which then may come to fit the original economic theory more closely (MacKenzie 2004). A moderate version of this performativity argues that the adoption of a theory can turn what was a relatively weak (or only valid in some conditions) to a stronger, more universal theory. A less moderate version argues that a theory can also be self-fulfilling by turning what was a false theory (at the time of theory development and introduction into the wider community), into being true. E.g., some argue that the field of economics' focus on the self-interest of economic actors leads students of economic theories to become more self-interested over time, thus constituting a self-fulfillment of the original theory (Frank and Schulze 2000; Ferraro, Pfeffer, and Sutton 2005). Finally, a concept less discussed in the literature is the potential self-falsifying nature of a theory (or counter-performativity cf. MacKenzie 2004), where a prophecy leads to behavior that makes the belief not true (Biggs 2009; Merton 1936). A recent study on racial bias among basketball referees shows that the racial bias disappeared after an academic study and media attention highlighted the problem (Pope, Price, and Wolfers 2013). The original study by Price and Wolfers (2010) could thus be argued to have been self-falsified.

Secondly, one can talk about *what is changed*, and different contributions refer to different concepts; self-fulfillment of theories, assumptions, worldviews, beliefs, prophecies, language and so forth (Ferraro, Pfeffer, and Sutton 2009, 2005; Felin and Foss 2009b, a; Merton 1948; MacKenzie 2006; Callon 1998; Bourdieu 1992; Biggs 2009). A worldview or a specific belief can clearly be influenced by a prophecy, e.g. if Warren Buffet declares a stock worthless (Merton 1948). But following the heated discussion in Ferraro et al. (2005; 2009) and Felin and Foss (2009a and 2009b), we are interested in whether and how the validity and integrity of social science *theories* (i.e. general mechanisms) – not just specific prophecies about isolated phenomena or general worldviews – are in question due to their potential self-fulfilling nature. Therefore, in the present paper we focus solely on the concept of self-fulfilling theories, in order to investigate whether social science's status as a science is threatened, and if the potential self-fulfilling nature of theories constitutes a particular ethical challenge.

2.2. Potential implications of social science theories

A theory will course not just be influential due to it being proposed, but requires some level of adoption in practice. Furthermore, the mere adoption of a theory does not necessarily imply the strong form of performativity. The following brief example illustrates how a theory can be adopted and be influential, without implying self-fulfillment; Daniel Kahneman has during his career offered numerous examples of how the thinking and behavior of human beings can be biased. In popular terms, system 1 is the type of thinking that is instantaneous and follows crude heuristics, while system 2 thinking is able to reflect upon these heuristics (Kahneman 2011). Kahneman is thus one of many who have enabled human beings to become aware of the limitations of system 1, an awareness that has the potential to change how humans interact in the social world. Proponents of social constructionism will argue that the influence of norms (how to think about biases), language (system 1 and system 2) and the role of institutions (teaching at universities and a Nobel Prize awarded to Kahneman in 2002) can facilitate a change in human behavior. Kahneman's theories are certainly performative in this generic sense. But while this behavioral theory might nudge the balance between system 1 and system 2, the theory will neither self-falsify nor self-fulfill itself, because finding the right balance between the systems is a key function of system 2. Changes in behavior are thus expected, and certainly in line with this type of theory about human behavior. While Kahneman's theory challenged how the field of economics should model human behavior (Hommes 2006), it is certainly no threat to social science as a science.

The archetypical example of self-fulfillment relates to the notion of the self-interested homo economicus (Ferraro et al. 2005). Various studies seem to show that exposure to economics changes how economic students behave in social science labs (Marwell and Ames 1981), how principles of fairness are integrated in responses to survey questions about economic issues (Frank, Gilovich, and Regan 1993) and the amount of money individuals might donate (Bauman and Rose 2011)³. E.g, being exposed to economics

³ The results are so far not entirely conclusive though, since some studies argue that the particular behavior of economic students compared with non-economic students might be due to selection effects, i.e. nature and not nurture (Cipriani et al. 2009).

makes students more inclined to accept that an owner of a retail shop raises the price of snow-shovels after a snow storm (Cipriani, Lubian, and Zago 2009) or how to balance layoffs and the financial performance of a firm (Rubinstein 2006). The argument is that students not only learn new economic concepts and tools, but also integrate new economic norms and perspectives that entail a persistent change in behavior. Since most economic models rely on (some number of) economic actors behaving rationally (i.e. self-interested), the change in behavior may make economic theories fit economic practice more directly, compared to an economic world where relatively fewer students are exposed to economic teaching. One might argue that self-interested behavior is more natural or more true economic behavior, but the theory would nevertheless still have changed typical behavior and thus be self-fulfilling, to some degree.

This example highlights the main issue at stake; not only will the behavior of individuals change to some degree due to the adoption of a scientific theory, but some underlying empirical regularities (Felin and Foss 2009a) will be changed also, which might turn the theories from false to true. Self-fulfillment is thus threatening to undermine objectivity in general and a fundamental premise of scientific realism in particular: That there are some sort of theoretical mechanisms (cf. Little 2015) and regularities out there in the world not directly affected by our theorizing about them. This issue constitutes the core of the recent discussion between management scholars Ferraro et al. (2005, 2009) and Felin and Foss (2009a, 2009b) and the discussion continues in accounting (Vosselman 2014), marketing (Mason, Kjellberg, and Hagberg 2015), finance (MacKenzie and Spears 2014) and economics (Esposito 2013). According to Felin and Foss (2009a, 2009b) the notion of self-interest is a constitutional part of human nature, and not a form of behavior that business schools have induced. Their crucial point is that if a form of behavior becomes more widespread after being made explicit in a theory, this is due to the theory being effective at capturing the underlying objective reality. In other words, if economic students become more self-interested after being exposed to economic theories about self-interest, it is simply because self-interest is

an effective type of economic behavior that can optimize economic outcomes – just like insight into chemistry can optimize how we handle chemical material.

We believe it is possible to reconcile these two different perspectives, both acknowledging the potential self-fulfilling change in behavior but also securing realism in social science. In order to cut through this discussion, we introduce different technical conceptions of realism and show that realism is not necessarily in opposition to changes in the (theorizing about) regularities of the external world.

3. Realism in natural vs. social science

Regarding the domain of natural science, scientific realism, briefly put is the view that science aims to provide us with true, or near true theories of the world, its properties and its structure. Realism is distinct from empiricism (and traditional formulations of positivism) in that it aims to describe the world beyond what is observable to us. In parallel, scientific realism regarding social science involves aiming at providing true theories about the social world that go beyond manifest phenomena (see Shearmur 1986 for a similar formulation). Realism is also distinct from various constructionist positions that deny the mind independent nature of the objects studied in the sciences. It is the discussion over the mind independent status and the ability of social science to potentially affect the objects of study and the theoretical mechanisms uncovered that is central to the debate over self-fulfilling theories.

Based on the above outline of scientific realism, scientific realism involves four different dimensions of realism: metaphysical realism, semantical realism, axiological realism and epistemological realism. Psillos (1999), Sankey (2001) and Niiniluoto (1999) take scientific realism (regarding natural science) to be a position comprised of all of these dimensions of realism. These dimensions can be combined in different ways, and some combinations imply a stronger version of realism, than others.

Metaphysical realism (MR): The world is objective and exists mind independently (Devitt 1984).

Realism about theories, or realism about theories in our context (RT): Scientific theories are either true or false independently of what we know (Hacking 1983: p. 27).

Axiological realism (AR): Science aims at the truth (Van Fraassen 1980).

Epistemic realism (ER): We are rationally warranted in believing that our best scientific theories are true (Psillos 1999).

Put in another way, the dimensions of scientific realism deal with the (objective) existence of objects, conceptions of truth, our epistemic access to assessing truth and the aim of science. Notice that Felin & Foss (2009b, p. 3) define their conception of scientific realism as: “...*there are objective facts and truths in the social world (independent of our theories about them)*”. They thus have a rather narrow, metaphysical (MR), conception of scientific realism in combination with a version of realism about theories (RT).

In putting the emphasis in this way, it is clear that scientific realism is contrasted with a position that questions the mind independency of the objects studied, and this is the main focus for Felin & Foss (cf. the above quote). As such there are clear affinities between scientific realist positions regarding natural science and social science. As we shall go on to suggest in the next section however, some care needs to be taken regarding the exact interpretation of MR, as its current formulation clearly favors an understanding of scientific realism relevant to natural science. Furthermore, in general it should be observed that a realist position can easily be had through a commitment to either all or a mixture of the different dimensions of realism suggested above, although we do not wish to deny that MR is central.

3.1. Realism in social science

Metaphysical realism (MR)

A fundamental characteristic of social science is the fact that social science theories are about human beings (who have minds) and/or socially constructed objects, like organizations and values that owe their existence to there existing human beings. MR requires that science is about mind independent objects,

and a mind independent object is defined as an object whose existence is not dependent on human beings or human experiences. Hence, nearly all of social science fails to qualify as metaphysical realism traditionally defined (when we consider scientific realism as formulated with regard to natural science), since social science focuses on humans and their socially constructed interactions. The above formulation of metaphysical realism is therefore inappropriate for capturing what realism amounts to with respect to social science.

A realist may proceed in the following way. In order for any discipline to count as scientific it has to satisfy certain conditions, and central among these conditions stands some form of objectivity. Different approaches to social science entail very different understandings of objectivity, but for our purposes we define this criterion of objectivity in the following way: a subject S of inquiry allows objective inquiry if it is possible for different inquirers to derive the same observations of S under the same conditions of observation (as one would commonly assume in social science experiments, e.g.). This is no more than a rough and ready definition of objectivity, but it will suffice for making our general point. It will also suffice for making the point that for large parts of subjects studied within social science by use of methods in the hermeneutic tradition the objectivity of data will be compromised, according to MR. In the case of investigating the extent to which self-fulfillment of theories may raise special considerations of concern for the claim to realism this is however irrelevant. We shall only concern ourselves with those areas that are potentially affected by self-fulfillment, because they are otherwise thought of as being the appropriate targets for a realist position.

The way that a mind independent subject of inquiry is supposed to differentiate itself from a mind dependent subject of inquiry is as follows. A mind independent subject of inquiry lends itself to objective inquiry, simply because it is thought of as being independent of the observer – and thus, an observer will not have any impact upon, or interfere with the subject in any way, so that other observers placed under the same or similar conditions would be able to obtain the same results as the original observer were they to put themselves in the position of the original observer. We should stress that this is a highly idealized

way of thinking about objectivity, and it ignores interesting cases where instruments for observation interferes with the phenomena to be studied (in the natural sciences a parallel example is sometimes mentioned in the study of quantum phenomena (Bohr 1958), but we will not consider that here). The point we wish to make here is simply that social phenomena, the subject of social science, are in no way different from mind independent subjects in this respect. The mind dependent status of the subject of inquiry is entirely indifferent to the issue of objectivity, simply because the mind dependency of the subject need not compromise objectivity when objectivity is understood as objectivity of inquiry.

What matters to realism in social science is whether it is possible to conduct studies in ways such that it is possible to get researcher independent results. In general therefore it is important to realize that epistemic objectivity and ontological objectivity can be related in the sense that a necessary condition upon something counting as epistemically objective is exactly that it's existence does not depend upon an inquiry having taken place – that the subject of study is not somehow an artificial construct resulting from an inquiry having taken place but not existing pre inquiry. In certain cases facts are dependend on an inquirer, e.g. when an ethnographer becomes part of a community in order to study this community (Malinowski 1944). But even in this field of ethnography, there is a worry about ‘going native’ and becoming too immersed in the community, since this implies a risk of “losing objectivity” (O'Reilly 2013: p. 88) – the same standard of objectivity and perspective on mind dependence thus still exists.

Nevertheless, discounting cases like this and focusing more narrowly on perspectives in social science where we are not interested in self-reports of this kind, an independent inquirer is taken to be the authoritative judge on the kinds of mind dependent facts that we are interested in describing. And he or she is independent exactly in the sense that he or she is studying mind dependent phenomena that in no way are dependent for their existence on the inquirer (in contrast to inquirees) in question. Panel data such as national statistics on salary and demographics would be an archetypical example. In this way, it is possible to conduct investigations on mind dependent phenomena, that are independent of the inquirer in an (epistemically) objective way. The prime example being the kinds of mechainsms postulated to exist

by scientific realists regarding social science, for the purpose of explaining social phenomena – ie. those unobservables taken to be responsible for the observable phenomena. So accepting that a traditional MR will not do, for capturing the obvious sense in which social phenomena are mind dependent in a general sense we need to reformulate this aspect of scientific realism.

Let us call a definition of metaphysical realism that emphasizes objectivity MRO.

MRO: Social phenomena exist mind dependently and can be studied objectively, i.e. independent of subjective influence on behalf of the person conducting the study.

In summary: Theories within social science are developed by scientists studying social phenomena (minds studying mind dependent phenomena). It is not the mind dependent nature of the subject *per se*, that poses a challenge to realism about social science. The issue is rather whether the phenomenon is dependent upon the mind of the observer. As long as the dependency relation is between groups or individuals, not identical with or in part constituted by the observer, objectivity need not be compromised.

While identifying the aspects of MR that are particular to social science, we have also discussed Felin & Foss's (2009b & 2009a) metaphysical notion of scientific realism. In the following, we shall attend to the other dimensions of realism in order to create a basis for examining whether these dimensions are in conflict with or are threatened by self-fulfillment. In this way, we can evaluate whether any version of realism is affected by the alleged threat of self-fulfillment.

3.2. Truth, success and social science

Axiological realism (AR), realism about theories (RT) and epistemic realism (ER)

Now consider how self-fulfillment could be an issue with regard to any of the remaining dimensions of realism. Self-fulfillment tells us that theories proposed at a time may be false, and yet successfully predict something in the future, in virtue of them being true at that time. This is certainly not in conflict with AR. Science may aim at the truth regardless of this. It is not in conflict with RT, because the change of truth value over time is not in conflict with theories being either true or false at times. Of course, if we did not

use tense, theories could come out both true and false, but that is not at issue, and there is nothing preventing us from using modal terms.

Is self-fulfillment supposed to be an issue for the epistemic component of realism? There is in fact a very straightforward way in which self-fulfillment may be considered to be a threat to ER but realising this requires us to focus on how realists argue for the conclusion that we are warranted in believing that theories are true. The most famous argument for scientific realism with regard to natural science is the so-called no miracles argument. This argument is supposed to motivate ER and RT in so far as the argument is that we have reasons to believe that our theories are true. The argument finds its canonical formulation in Putnam (1975: p. 73) who claims that realism “*is the only philosophy that doesn't make the success of science a miracle*”. Scientific theories are assumed to be successful, and what drives this success? One explanation is that our best theories are true (or approximately true). Indeed, if these theories were far from the truth, so the argument goes, the fact that they are so successful would be miraculous - for elaborations of the miracle argument, see Brown (1982), Psillos (1999), Lyons (2003) and Busch (2008).

The important point here is that realists require a reliable link between truth and predictive success. Now, what self-fulfillment supposedly does is to undermine this connection. A scientific realist will argue that it is reasonable to believe that theories are true when they are predictive. However if some of these theories have in fact started off their lives as being false, the reliable link between truth and success has been severed. The potential self-fulfillment of theories thus appears to undermine a central argument for realism, the argument motivating ER. In the following sections, we will introduce new theoretical perspectives to the discussion, in order to show why none of the four elements of realism are compromised by self-fulfillment.

4. A realistic social science realism: Cooling the fire

4.1. Time operators, under-determination and an ever-changing world

The notion of time has been overlooked in the discussion on self-fulfillment so far, and the role of time operators hence needs to be presented in detail, in order to show what a proper defense of scientific realism will look like. Because truth is supposed to explain success, it is important to point out that theories are only considered to be successful at the time when their predictions match reality (regardless of whether we are considering self-fulfilling theories or not). So if a theory is factually false prior to the time of its introduction (T^1) but is confirmed to be true at the time of testing it (T^2) – its success would still be explainable in terms of its truth. Theories make modal claims. The theory was not true prior to the time of creation (T^1) but presumably, the theory is not believed to be true at that time because it is supposed to be predictively (T^2) successful. For any realists operating with such a notion of success tightly correlated with predictive success (i.e. after the theory has been tested) no problem really exists.

The intention of the ‘time operator’ argument is to show, that if self-fulfilling theories are successful when they are true, pointing out that a theory is self-fulfilling is supposedly not a problem for a realist. However, since we only discover that theories are self-fulfilling in retrospect, a very different challenge could arise and another perspective needs to be introduced. If there is good evidence that self-fulfillment is a possible phenomenon (as is assumed by Ferraro et al. (2005 & 2009) and contested by Felin and Foss (2009a & 2009b)), and self-fulfilling vs. non-self-fulfilling theories are indistinguishable from the point of view of scientists previous to it having been established if they are one or the other, then it seems we never have good evidence for believing theories prior to their introduction. It thus seems as if self-fulfillment may imply a unique⁴ kind of under-determination of theory by data argument. Even if theories are in most cases not self-fulfilling and successful, and a realist would in those circumstances be warranted in inferring that those theories were true prior to their introduction, the mere presence of self-

⁴ Usually under-determination implies that the available data is insufficient for supporting one theory consistent with the data over another theory also consistent with the data, and in principle new data can solve the problem (Boyd 1973). In the case of self-fulfillment, it is not possible to get different data to solve the under-determination.

fulfilling theories undermines this inference. It is indeed under-determined by the evidence at a later time if theories were true or false at an earlier time. The phenomenon of self-fulfillment thus has the consequence that we are not necessarily entitled to say, upon having confirmed a theory, that the theory was true prior to the time of introduction. But this should not necessarily worry a realist. After all, theories in social science are not expected to be true in all domains at all times. Human nature and social behavior is, at least to some degree and within boundary conditions (cf. Felin & Foss 2009b), malleable, across domains and time and thus, a theory being true at one time may turn out to be false at a later time. E.g., the level of self-interested behavior is likely to change over time, whether or not such a change is correlated with the development of theories on the nature and efficiency of self-interest.

One might be worried about this potential for change of regularities if one requires the mechanisms explained and predicted by social science theories to constitute Newtonian, eternal laws (Longo, Montévil, and Kauffman 2012; Koppl et al. 2015). Our point here is that there is nothing adverse to social science realism, since this position does not require the truth of a non - dynamic theory or the existence of such unmalleable laws. Felin & Foss (2009a & 2009b) and Ferraro et al. (2005 & 2009) clearly disagree on the empirical scope of performativity broadly understood (not just self-fulfillment), but that is an empirical question, not an issue about the analytical nature of self-fulfillment and social science theories. The empirical scope has an impact on the overall integrity of the scientific community though, an issue to be addressed in section 5.

4.2. Meta-theories

Even if a social scientist does not accept the above arguments, and insists on a strong version of scientific realism and a demand for clear and predictable regularities and theoretical mechanisms there is an even stronger, more intuitive and very simple argument for why self-fulfillment does not constitute a threat to realism. A meta-theory is here defined as a theory that takes theories about empirical phenomena as its subject of explanation. A meta-theory is thus a second-order theory, concerned with the development of

first-order theories and their impact⁵. First-order theories are supposed to be able to explain underlying mechanisms and (probabilistically) predict human behavior. Usually such theories examine the impact of X on Y, e.g. the impact of economic teaching on self-interested behavior. Now, what if we carried out an empirical study of potentially ‘self-fulfilling’ theories? (as Ferraro et al. (2009) and Santos and Rodrigues (2009) call for). Following our conceptualization, one can develop meta-theories on whether the ‘impact of economic teaching on self-interested behavior’, has any effect on the ‘application of economic models’. Such a study would be similar to a study of how new economic tools impact financial markets (e.g. the impact of the Black-Scholes-Merton formula (MacKenzie 2006) or the Sunshine act (MacKenzie 2004). In this way it should, at least theoretically, be possible to create a (second-order) meta-theory that not only a) predicts if there is a self-fulfilling impact of the adoptions of (first-order) theories, but also b) identifies the specific theoretical mechanisms that constitute this impact. E.g., Price and Wolfers’ (2010) study on racial bias among basketball referees could have incorporated a (second-order) analysis of how the publication of this first-order theory can lead to new regulations, different training of umpires and just generally an awareness of racial bias – changes that could lead to a change of behavior and explain the self-falsification of their own original study in this particular setting. In the case of the argued change in self-interested behavior among economics student, one could identify whether it is the syllabus used, the teachers’ beliefs or other elements that influences behavior (Cokgezen and Cilingirturk 2014). Even if one rejects the existence of self-fulfillment, but ‘merely’ accepts simple, generic performativity, one acknowledges that (the adoption of) theories can influence social behavior, and in order to properly assess theories and the impact adoption might have, one should actually carry out such comparative, social science experiments, as pointed out by Santos (2011).

⁵ We do not therefore intend the term ‘meta’ to indicate that we are concerned with the methodology or reliability of theories.

It is not important to our analytical point about self-fulfillment and the threat to realism that the sense in which this is ‘possible’ may be fairly remote in terms of our actual epistemic situation⁶. But, our argument is premised on it being possible in principle to conduct such a study, i.e. there is nothing in the nature of the phenomena to be theorized about that prohibits such an investigation.

This crucial, yet simple, point has been overlooked in the previous debate in this field. No realist of any kind should have any problems with a situation where a first-order theory changes people’s behavior, if this change can be predicted by a meta-theory – and being a realist, the realist would merely need to argue that this is possible.

5. Ethics and intentionality: Implications

Some researchers are worried about the potential detrimental effect that performativity in general and self-fulfillment specifically might have. Ferraro et al. (2005 & 2009) argue for example that evidence shows that people educated in departments of economics or business schools are ethically compromised as a result of having been exposed to theories in these domains, meaning that they act out of self-interest to a higher degree than students from other subject areas (cf. footnote 2, it still is not clear whether the effect is primarily due to selection or learning effects). Such examples seem to indicate that self-fulfillment raises a unique kind of ethical challenge and Ferraro et al. (2009: p. 673) explicitly claim that: “...we need to take responsibility for the consequences of what we teach and write”. Hence, the argument is that we, academics, need to be careful as to what we theorize about and what conclusions we reach, as the outcome may have ethical implications in virtue of the self-fulfilling nature of theories in these areas (Ghoshal 2005; Ferraro, Pfeffer, and Sutton 2009). But, in what sense can a scientist be held ethically accountable for the effects of the potential self-fulfilling nature of their theories?

A key issue is the assumption one makes about scientists and their (moral) aims: Are they trying to be as truthful as possible, or attempting to distribute lies? If they are distributing lies, they are naturally

⁶ Similarly, the ‘three-body problem’ constitutes an epistemic challenge that prohibits a clear analytical solution (Strogatz 2011).

ethically accountable, no matter whether the lie turns out to be true or not. In the following we therefore rely on the assumption that scientists aim to present what they think are true theories⁷. In order for self-fulfillment to imply a unique kind of ethical responsibility, the researcher must be in a position to know that the particular theory being suggested will establish changes to the (at the time of theory development) current state of the world. Hence, the researcher has to know of the impact of the given theory, and if he knows the impact, it becomes questionable to what extent he knows that the theory may be self-fulfilling if he satisfies the requirement about not consciously disseminating false theories. You can only be moral (or immoral) if you know or even suspect the (moral) outcome (Williams 1985). Furthermore, as argued in the previous sections, if a given researcher knows the impact of the theory, he basically has a meta-theory that predicts a possible future state of the world, which again means he has not changed the world from true to false. On the other hand, if the researcher is unable to predict what kind of impact his theory will have, any outcomes are unintentional and it is difficult to ascribe moral value to the actions. Hence it is not self-fulfillment specifically, but performativity in general that implies an ethical accountability. But the arguments concerning the potential ethical accountability of proposing theories should not just hinge on an individual researcher and theory level, i.e. be focused on the outcome of a specific theory. We argue that such a level does not make explicit that the ethical responsibility not just relates to the a) academic theory-building activities but it might make more sense to talk of a b) community-level responsibility.

As Ferraro et al. (2005 & 2009) have argued, the institutional settings are influential and some of these institutional features are outside of the theory-building realm. Therefore, as in the natural sciences, it's not just theories per se that have ethical implications, but also how they are applied and in what setting. The Black-Scholes-Merton example illustrates this clearly. The academics involved did not merely propose a model, but were entrepreneurial and deliberately tried to shape the institutional settings of the financial world (MacKenzie 2006; Haug and Taleb 2011). In this sense, they went beyond the strictly theory-

⁷ While the assumption might seem strong given the various biases that researchers have been shown to have (Fanelli 2010; Francis 2012) it is relevant to point out that Merton (1973) did not explicitly include honesty in his conceptualisation of the virtues of science, which thus arguably was taken for granted. We thank Kieran Healy for this point.

building aims. Hence, if one argues that the model was self-fulfilling, or even just performative in a more general sense, the analytical treatment of this issue shows that any ethical responsibility does not just relate to proposing the theoretical model but on activities going beyond the mere academic realm.

It should thus be clear that it is not self-fulfillment per se that implies an ethical responsibility. The overall performative nature of social science clearly shows though, that theorizing matters and researchers should reflect on the potential impact of their activities. If one relies on the assumption that people are self-interested (here considered to be an assumption of a scientific community), one needs to reflect on the potential consequences of such an assumption for theorizing. Given the complexity of social science, this is a significant challenge. To predict the extent to which the norms and language of a science entrenches and instantiates social practices is at the very least extremely difficult, since such changes not only go from strictly false to strictly true, but have much more blurry boundaries, being true or false to some extent in some contexts. This more general challenge is beyond the scope of this paper though, where we have focused on cooling the fire in the debate on self-fulfillment.

6. Conclusion

In the present paper, we have dismantled the dual threat of self-fulfillment by offering an analytical examination of how self-fulfillment relates to scientific realism and ethics. Hereby we have tried to cool the fire, by explaining how self-fulfillment is not a particular threat to scientific realism nor any constructionist conception of science and ethics.

An extensive presentation of different dimensions of realism illustrates how the previous debate on this issue reveals a number of substantial misunderstandings. First of all, the social world can change (possibly merely within boundary conditions), without undermining realism and hence objectivity. Secondly, realism as a philosophy of science position is saved via the concept of a meta-theory. Any realist should accept a changing world, if one can produce a meta-theory that predicts the self-fulfillment by explaining the underlying mechanisms that have caused the change in the social world. Therefore, self-

fulfillment is not, analytically speaking, a fundamental threat to a social science realism, as has been assumed by writers such as Ferraro et al. (2005 & 2009) and Felin & Foss (2009b).

In terms of the ethical challenge, a researcher can only be held (morally) responsible, if he knows (or suspects) the (moral) outcome. If he knows the outcome of the introduction of a given theory, he basically has a meta-theory. Thus, the researchers can not be held morally responsible for the potential self-fulfillment of a given theory. Any responsibility is thus due to the general impact of performativity, and not self-fulfillment in particular. Finally, we argue that since one can not trace the impact of an individual theory, it makes more sense to talk of a general responsibility for the world-view of one's scientific community.

7. References

- Aronson, Joshua, Carrie B. Fried, and Catherine Good. 2002. "Reducing the Effects of Stereotype Threat on African American College Students by Shaping Theories of Intelligence." *Journal of Experimental Social Psychology* no. 38 (2):113-125.
- Bauman, Yoram, and Elaina Rose. 2011. "Selection or indoctrination: Why do economic students donate less than the rest?" *Journal of Economic Behavior & Organization* no. 79:318-327.
- Biggs, Michael. 2009. "Self-fulfilling prophecies." In *The Oxford Handbook of Analytical Sociology*, edited by Peter Hedström and Peter Bearman. Oxford: Oxford University Press.
- Boghossian, Peter. 2006. *Fear of Knowledge: Against Relativism and Constructivism*. Oxford University Press.
- Bohr, Niels. 1958. *Atomic Physics and Human Knowledge*. New York: John Wiley & Sons.
- Bourdieu, Pierre. 1992. *Invitation to a Reflexive Sociology*. Chicago: University of Chicago Press.
- Boyd, Richard N. 1973. "Realism, underdetermination, and a causal theory of evidence." *Nous* no. 7 (1):1-12.
- Brown, James Robert. 1982. "The Miracle of Science." *Philosophical Quarterly* no. 32:232-244.
- Busch, Jacob. 2008. "No New Miracles, Same Old Tricks." *Theoria* no. 74:102-114.
- Callon, Michael. 1998. "Introduction: The embeddedness of economic markets in economics." In *The Laws of the Market*, edited by Michael Callon, 311-357. Oxford: Blackwell Publishers.
- Cipriani, Giam Pietro, Diego Lubian, and Angelo Zago. 2009. "Natural born economists?" *Journal of Economic Psychology* no. 30:455-468.
- Cokgezen, Murat, and Ahmet Mete Cilingirturk. 2014. "Attitudes of Turkish students against markets: does studying economics matter?" *International Journal of Social Economics* no. 41 (10):944-955.
- Devitt, Michael. 1984. *Realism and Truth*. Princeton: Princeton University Press.
- Esposito, Elena. 2013. "The structures of uncertainty: performativity and unpredictability in economic operations." *Economy and Society* no. 42 (1):102-129.
- Fanelli, Daniele. 2010. "Do Pressures to Publish Increase Scientists' Bias? An Empirical Support from US States Data." *PLoS ONE* no. 5 (4).

- Felin, Teppo, and Nicolai J. Foss. 2009a. "Performativity of Theory, Arbitrary Conventions, and Possible Worlds: A Reality Check." *Organization Science* no. 20 (3):676-678.
- Felin, Teppo, and Nicolai J. Foss. 2009b. "Social Reality, the Boundaries of Self-Fulfilling Prophecy, and Economics." *Organization Science* no. 20 (3):654-668.
- Ferraro, Fabrizio, Jeffrey Pfeffer, and Robert I. Sutton. 2005. "Economics Language and Assumptions: How Theories can Become Self-fulfilling." *Academy of Management Review* no. 30 (1):8-24.
- Ferraro, Fabrizio, Jeffrey Pfeffer, and Robert I. Sutton. 2009. "How and Why Theories Matter: A Comment on Felin and Foss (2009)." *Organization Science* no. 20 (3):669-675.
- Francis, Gregory. 2012. "Too good to be true: Publication bias in two prominent studies from experimental psychology." *Psychonomic Bulletin & Review* no. 19 (2):151-156.
- Frank, Björn, and Günther G. Schulze. 2000. "Does economics make citizens corrupt?" *Journal of Economic Behavior & Organization* no. 43 (1):101-113.
- Frank, Robert H., Thomas Gilovich, and Dennis T. Regan. 1993. "Does Studying Economics Inhibit Cooperation?" *The Journal of Economic Perspectives* no. 7 (2):159-171.
- Ghoshal, Sumantra. 2005. "Bad management theories are destroying good management practices." *Academy of Management Learning & Education* no. 4:75-91.
- Hacking, Ian. 1983. *Representing and Intervening*. New York: Cambridge University Press.
- Haug, Espen Gaarder, and Nassim Nicholas Taleb. 2011. "Option traders use (very) sophisticated heuristics, never the Black-Scholes-Merton formula." *Journal of Economic Behavior & Organization* no. 77:97-106.
- Hommes, Cars H. 2006. "Heterogeneous agent models in economics and finance." In *Handbook of computational economics*, edited by Karl Schmedders and Kenneth L. Judd, 1109-1186. Elsevier.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. London: Penguin.
- Koppl, Roger, Stuart Kauffman, Teppo Felin, and Giuseppe Longo. 2015. "Economics for a creative world." *Journal of Institutional Economics* no. 11 (1):1-31.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory Life: the Social Construction of Scientific Facts*. Princeton, New Jersey: Princeton University Press.
- Little, Daniel. 2015. "Mechanisms and Method." *Philosophy of the Social Sciences* no. Forthcoming.
- Longo, Giuseppe, Maël Montévil, and Stuart Kauffman. 2012. No entailing laws, but enablement in the evolution of the biosphere. Paper read at 14th International Conference on Genetic and Evolutionary Computation Conference.
- Lyons, Timothy D. 2003. "Explaining the Success of a Scientific Theory." *Philosophy of Science* no. 70:891-901.
- MacKenzie, Daniel. 2006. *An Engine, Not a Camera: How Financial Models Shape Markets*. Cambridge, MA: MIT Press.
- MacKenzie, Donald. 2004. "The big, bad wolf and the rational market: portfolio insurance, the 1987 crash and the performativity of economics." *Economy and Society* no. 33 (3):303-334.
- MacKenzie, Donald, and Taylor Spears. 2014. "'The formula that killed Wall Street': The Gaussian copula and modelling practices in investment banking." *Social Studies of Science* no. 44 (3):393-417.
- Malinowski, Bronislaw. 1944. *A Scientific Theory of Culture and other Essays*. Durham: The University of North Carolina Press.
- Marwell, Gerald, and Ruth E. Ames. 1981. "Economists free ride, does anyone else?" *Journal of Public Economics* no. 15:295-310.
- Mason, Katy, Hans Kjellberg, and Johan Hagberg. 2015. "Exploring the performativity of marketing: theories, practices and devices." *Journal of Marketing Management* no. 31 (1-2):1-15.
- Merton, Robert K. 1936. "The Unanticipated Consequences of Purposive Social Action." *American Sociological Review* no. 1 (894-904).
- Merton, Robert K. 1948. "The self-fulfilling prophecy." *Antioch Rev* no. 8:193-210.

- Merton, Robert K. 1973. "The Normative Structure of Science." In *The Sociology of Science: Theoretical and Empirical Investigations*, edited by Robert K. Merton. Chicago: University of Chicago Press.
- Niiniluoto, Ilkka. 1999. *Critical Scientific Realism*. Oxford: Oxford University Press.
- O'Reilly, Karen. 2013. *Key Concepts in Ethnography*: Sage.
- Pope, Devin G., Joseph Price, and Justin Wolfers. 2013. Awareness Reduces Racial Bias. In *NBER*.
- Price, Joseph, and Justin Wolfers. 2010. "Racial Discrimination Among NBA Referees." *The Quarterly Journal of Economics* no. 125 (4):1859-1887.
- Psillos, Stathis. 1999. *Scientific Realism: How science tracks the truth*. London: Routledge.
- Putnam, Hilary. 1975. "Mind, Language and Reality." In *Philosophical Papers, vol. 2*. Cambridge: Cambridge University Press.
- Rubinstein, Ariel. 2006. "A Sceptic's Comment on the Study of Economics." *The Economic Journal* no. 116 (510):C1-C9.
- Sankey, Howard. 2001. "Scientific Realism: An Elaboration and A Defence." *Theoria* no. 98:35-54.
- Santos, Ana C. 2011. "Behavioural and experimental economics: are they really transforming economics?" *Cambridge Journal of Economics* no. 35:705-728.
- Santos, Ana C., and Joao Rodrigues. 2009. "Economics as social engineering? Questioning the performativity thesis." *Cambridge Journal of Economics* no. 33:985-1000.
- Shearmur, Jeremy. 1986. "Realism Under Attack?" *Phil. Soc. Sci.* no. 16:219-222.
- Sherif, Muzafer. 1937. "An Experimental Approach to the Study of Attitudes." *Sociometry* no. 1 (1/2):90-98.
- Strogatz, Steven H. 2001. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering (Studies in Nonlinearity)*: Westview Press.
- Taylor, Charles. 1983. "Social Theory as Practice." In *Philosophy and the Human Sciences*, edited by Charles Taylor. Cambridge: Cambridge University Press. Original edition, *Social Theory as Practice*, The B.N. Ganguli Memorial Lectures 1981, Delhi: Oxford University Press, 1983.
- Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press.
- Vosselman, Ed. 2014. "The 'performativity thesis' and its critics: Towards a relational ontology of management accounting." *Accounting and Business Research* no. 44 (2):181-203.
- Williams, Bernard. 1985. *Ethics and the limits of philosophy*. London: Fontana.