How and Why Theories Matter: A Comment on Felin and Foss (2009)

Fabrizio Ferraro
IESE Business School, 08034 Barcelona, Spain, fferraro@iese.edu

Jeffrey Pfeffer
Graduate School of Business, Stanford University, Stanford, California 94305, pfeffer_jeffrey@gsb.stanford.edu

Robert I. Sutton
Department of Management Science and Engineering, Stanford University, Stanford, California 94305, bobsut@stanford.edu

Theories matter because they affect behavior and can, under certain circumstances, become self-fulfilling. For a theory to become self-fulfilling, people must be aware of the theory and have the ability to make choices according to its dictates, social and physical arrangements are altered on the basis of the theory’s prescriptions, and the proponents have the power to implement social arrangements consistent with the theory. Economics and other social science theories often fulfill these conditions, with implications not only for the work of scholars, but also for how we think about testing theories that can change the world they describe.

Key words: self-fulfilling prophecies; performativity; sociology of science

History: Published online in Articles in Advance April 27, 2009.

Since we can never know anything for sure, it is simply not worth searching for certainty; but it is well worth searching for truth; and we do this chiefly by searching for mistakes, so that we can correct them. (Popper 1992, p. 4)

We are honored and flattered that Felin and Foss (2009) elected to focus such a long critique on the ideas and arguments in our 2005 Academy of Management Review article. Nevertheless, we probably do not deserve such undivided attention because the mechanisms through which theories can become self-confirming have been the focus of theoretical and empirical analysis by diverse scholars from a spectrum of academic fields. In sociology, the performativity perspective, to use Callon’s (1998) terminology, is gaining momentum (see, for instance, MacKenzie 2006, MacKinnon 2006, MacKenzie et al. 2007, Callon et al. 2007, Espeland and Sauder 2007, Biggs 2009) as a theoretical lens to understand modernity and markets (Fourcade 2007, Fliigstein and Dauter 2008). In political science and political sociology, ideational scholarship has helped to better explain some of the most important historical transformations in societies (Blyth 2002, Somers and Block 2005). And in economics, there is an important area of theoretical development dealing with subjective rationality and self-confirming equilibria that considers the interrelationships among beliefs, actions, and outcomes (Kalai and Lehrer 1995, Ryall 2003; see also Arce (2007) for a model of self-activating agency dynamics).

Given the momentum surrounding these ideas, critique and analysis is welcome because we believe that articulating and testing multiple, competing theoretical perspectives is essential for advancing our knowledge. Unfortunately, Felin and Foss (2009) (hereinafter, F&F) do not seriously engage with the three mechanisms (institutional design, norms, and language) and the two scope conditions (culture and accountability) that we suggested can help explain how theories become self-fulfilling. Instead, they argue that because (1) reality is objective and (2) human nature is not very malleable, therefore (3) self-fulfilling prophecies do not exist or are extremely rare, and theories that are not true can not become true by changing the conditions in the world. Although F&F suggest that they are providing scope conditions to our theory, if taken seriously, their scope conditions would take us back to square one: only true theories can affect the social world, and self-fulfilling prophecies are just noise, possibly appearing for some short time but in the end succumbing to the “facts” of human nature and objective reality. They also take issue with our (and many others’) characterization of the sometimes harmful effects of the assumptions and language of economic theory. We consider each of these issues in turn.
Not Every Theory Can Be Self-Confirming, But Some Can Be

F&F both implicitly and explicitly argue that the theories that succeed are the ones that are most veridical with the world as it exists. This position, at least in its extreme form, clashes with the large literature on the sociology of science, beginning with Kuhn (1970). In addressing the question of how and when new theories or paradigms replace an existing order, Kuhn (1970) argued that disconfirming evidence alone was seldom sufficient to dislodge a paradigmatic consensus. The implications of his work and other, empirical work in the Kuhnian tradition (e.g., Lodahl and Gordon 1972) is that the acceptance of scientific theories is a function of both their truth value and the political and rhetorical skills and power of their proponents and opponents (see, for example, Morrell (2008) for an articulation of this viewpoint with respect to evidence-based management).

It is well beyond the scope of this short rejoinder to review the vast literature in the philosophy and sociology of science that belies the claim that theories succeed or fail solely on the basis of their “truth” value.1 It is worth noting how F&F, despite their declared allegiance to critical realism (Haack 1998), end up with arguments that are much more akin to 19th century logical positivism. Even Popper (1959) would have to take issue with F&F from an epistemological point of view, because Popper assumed that some theories are true, but we will not necessarily know which ones.2

With their recourse to ideas such as fundamental human nature, which they seem to believe exists apart from things such as social influence and perhaps even learning, F&F engage the “nature” versus “nurture” debate. The relevant literature concerning this argument is truly vast and includes most of social psychology, sociology, political science, economics, and numerous other fields as well. Suffice it to say for our purposes that if human nature were truly fixed and immutable, there would be no reason to teach students economics, because such learning could not alter their “human nature” and therefore how they make choices.

Learning, both at the individual and societal levels, helps us explore the limits of our individual and societal possibilities, and although we would (ontologically) agree with F&F that these limits are possibly finite, their argument seems to imply that we actually know what these limits are, whereas we argue that we cannot know them (an epistemological difference). Indeed our position would be consistent with an open system approach (e.g., Miller 1965) and with the idea of equifinality, according to which “a system can reach the same final state, from different initial conditions and by a variety of paths” (Katz and Kahn 1978, p. 30). We cannot know, ex ante, all of the feasible paths, and we try to do our best, “searching for truth,” to use Popper’s (1992) words; but the possible paths are only created on the basis of our theorizing, which is more akin to a flashlight in a dark room than to a set of engineering blueprints.

It is also important to reemphasize that although the focus of our analysis was primarily economic theory, it was not exclusively so. Our argument is that many social science theories, including theories in psychology (Schwartz 1997) and sociology, can be self-confirming. Dweck’s (2006) extensive research, for example, shows that people who believe that their IQ is fixed are more likely to avoid situations where they are likely to fail, less likely to persist in their efforts in the face of failure, and therefore less likely to learn new skills when compared to people who believe that their IQs are malleable. Dweck (2006) also provides evidence that these beliefs about IQ can be changed through a social influence process, providing still more support for the idea that theories can be self-fulfilling and that human nature is neither fixed nor immutable (thereby undermining two of F&F’s three main assertions). Consistent with Dweck’s (2006) arguments, a longitudinal study by Aronson et al. (2002) showed that African-American students at Stanford University who were persuaded that their IQs could be changed (rather than are fixed) earned higher grades during the next term.

F&F, closely and carefully read, discount the idea of the self-fulfilling prophecy in that they argue that for a self-fulfilling prophecy to become true, it must be actually correct. But there is strong empirical evidence that in many instances expectations and prophecies do become true because they cause alterations of structures and processes that affect the world in self-fulfilling ways. We would not deny that not all instances of self-fulfilling effects have been empirically confirmed (e.g., Brophy 1983) or that self-fulfilling expectations affect everyone equally at all times (e.g., Madon et al. 1997, Jussim and Harber 2005; see also the fascinating field experiment conducted by Salganik and Watts (2008)). It is, of course, patently the case that not every theory can become self-fulfilling, so in that sense, F&F are quite correct when they speak about the need to specify the boundary conditions for the operation of self-fulfilling theories. Yet their article fails to specify any such scope conditions (except to talk about whether or not theories are actually true as a scope condition), so we do so here.

For a theory to have any potential to become self-fulfilling, several conditions must hold. First and most importantly, the subjects of the theorizing must have the ability to both apprehend the theory and make choices and take actions based on the theory’s dictates. As Callon (2007, p. 322) noted, a “model can be self-fulfilling because it is all about the behaviors of human beings, and human beings depend on beliefs and expectations that planets do not have.” This scope condition means that self-fulfilling theories almost certainly don’t apply to the movement of planetary bodies, chemical
reactions, and other physical phenomena that are concerned with inanimate objects that are not going to be affected by the theory. And for a theory to be self-fulfilling, the objects of the theory must know about it so they can potentially act on its dictates—which implies both literacy of and familiarity with the theory, as well as some elapsed time between knowing the theory and making predictions based on it during which decisions and actions can occur (Henshel 1982). One reason that economic ideas are so likely to be self-fulfilling compared to other social science theories is that economic concepts are more pervasive and dominant in everyday discourse and understanding.

What is definitely not needed for self-fulfilling prophecies and theories to become true, contrary to what F&F argue, is the assumption that individuals are irrational or readily duped—something that we never implied in the first place. Models of self-fulfilling prophecies can and have been built with rational action models of decision making in competitive settings. Kalai and Lehrer (1995), for instance, showed that in an oligopoly, if managers believe they are in a competitive market, the (subjectively) rational response is to produce the competitive level of output. But this decision results in market prices equal to marginal costs, and therefore an industry can become highly competitive if participants believe it is so, even if the initial structural conditions were different. This class of model can also help explain how competitive advantage can be sustained (Ryall 2003; see also Brandenburger (2007) for a review of the epistemic program in game theory, which offers game theory a formal language to model participants’ beliefs and knowledge).

In considering the scope conditions for self-confirming theories, however, even in the physical realm, some caution about the extent to which theories can not alter the objects of their study may be in order. For instance, in physics, the Heisenberg uncertainty principle and the observer effect both hold that measurement of even physical objects can affect their momentum, and vice versa. In medicine, a science that concerns itself with biological processes that are presumably both physically real and immutable, the placebo effect is nevertheless so potent that double-blind studies have become the gold standard for drug approval. Apparently, even human biology can, at least to some extent, succumb to expectations effects.

For theories to become self-fulfilling, it is not just that people must believe them and act on those beliefs, although these are necessary conditions. For a theory to become true by construction, social and even physical arrangements must be altered on the basis of the theory in ways that make it likely to become true. Such changes in arrangements can entail alterations in management practices and organizational and even physical design. Note that in the case of the Black-Scholes option pricing model (MacKenzie and Millo 2003), the model’s “accuracy” increased when “cheat sheets” using the model to calculate prices were distributed and when the formula was incorporated into software made available to Chicago Board Options Exchange market participants. Beliefs that are incorporated into actual technology, practices, and social arrangements are more likely to become self-fulfilling and to affect the world in a self-confirming way over time. In the case of options pricing, Fischer Black was not just a disinterested academic, but an active promoter of the formula—an idea entrepreneur. In fact, as Beunza (2008, p. 95) noted, “the emerging lessons of Black-Scholes to budding performateurs are no different from the usual practice among purveyors of technological standards: keep it simple, pitch high, sell low… and throw in the accessories.”

F&F’s attempt to discount the self-fulfilling aspect of the implementation of the Black-Scholes option pricing model on the Chicago Options Exchange is misleading and does not do justice to the careful historical work of MacKenzie and Millo (2003, see also MacKenzie 2006, 2007). Contrary to F&F’s assertion, the Black-Scholes model did not triumph because it was somehow more correct and superior to alternative ways of pricing options. When the model was first introduced, there were wide discrepancies between its predictions and preexisting price patterns. The Black-Scholes model was not, ex ante, a great representation of reality. It was instead widely criticized for its unrealistic assumptions (Gastineau 1975). The model, though, was mathematically elegant, had high academic standing, and was, despite its sophistication, remarkably user friendly because it required traders to understand and discuss only one parameter—implied volatility. Most of the alternative models required estimation of at least three or more parameters. After the U.S. stock market fell on October 19, 1987, the ability of the Black-Scholes model to predict market prices has decreased significantly (MacKenzie 2006).

It is also interesting to note what F&F use as an example of what is not self-fulfilling from economics: hyperrationality or agent omniscience. Simply knowing about hyperrationality can not in and of itself make one able to achieve greater cognitive scope, although such knowledge and belief in the advantages of hyperrationality might provide incentives to implement technologies that would make the condition more likely to be achieved. But the fact that a theory (and its acceptance) can not by itself enable an individual to achieve full knowledge of all preferences and choices does not logically mean that such a theory is unable to influence whether or not people choose to behave in a self-interested fashion or to decide to choose profit maximization over other values in decision dilemmas, things that are within their decision scope and possibility.

Implied in the second scope condition above is another that we ought to make explicit: the power to actually implement the changes in organizational arrangements
and management practices that can make a theory true by construction. In this regard, there is again an important lesson for the other social sciences from the history of economics. In the late 1920s, about one-third of the membership of the American Economic Association (AEA) signed a statement opposing the passage of the Smoot-Hawley tariff, a bill that when implemented was widely believed to have made the Great Depression worse. When the economists’ pleas were ignored by President Hoover, who signed the bill, the economists and their professional association, the AEA, vowed to not be so readily ignored again. The history of how the field of economics accomplished its increase in influence (Bernstein 2001) provides a lesson for other social sciences. But it also reminds us that a third scope condition for theories to become self-fulfilling is the power of their proponents, for it requires social power to implement the technologies and other organizational arrangements necessary for theories to become true. The activism of economists during the recent global financial crisis, and the media attention their “open letter” received, provided yet more evidence of the discipline’s ability and willingness to vigorously play the policy game and of the inability (or unwillingness) of other social sciences to enter this arena.

Effects of Studying Economics and Business

In addition to arguing that theories can not become self-fulfilling, F&F argue that economics is harmless at worst or, more likely, beneficial, with its emphasis on self-interest, strong incentives, and individual choice. Our principal point was that theories can be self-confirming and, therefore, theories matter. There are certainly parts of economic theory that can lead to what at least some might consider socially undesirable choices and behaviors, but there are obviously also socially useful and constructive parts of economic theory as well. We note just a few things here.

First, many of the “experiments” F&F discount in their article are, in fact, precisely the same types of experiments, including those using the prisoner’s dilemma paradigm, that are part and parcel of the agenda of experimental economics. We presume that because many of the studies in the experimental economics tradition show that incentives operate as predicted by standard economic theory, F&F would be happy to accept those results. Denigrating the ability of experiments to provide insight into human behavior not only all too casually dismisses the contributions of experimental economics—and we note that many of the results we cite were, in fact, published in economics journals—it also basically disregards nearly all of social psychology and a good portion of sociology as well.

Second, there is quite a bit of evidence on the harmful effects of economics and business training on ethics and ethical behavior that does not rely just on the laboratory. In studies of cheating, McCabe et al. (2006) reported that graduate business students cheat more than their non-business school peers. McCabe (1997) reported similar differences for undergraduates, with business school students cheating more. McCabe et al. (1991) found that business students made more unethical choices in decision dilemmas than did law students. Frank and Schulze (2000) found that in a German university, economics students were more corruptible in that economics students were more likely to recommend a plumber for a film club that charged a higher price when they received more money for making such a recommendation.

More recent evidence from a field experiment in Switzerland highlighted the effects of self-selection, as business students contributed less to funds for needy students from the very beginning of their business studies (Frey and Meier 2003). Nevertheless, even if self-selection into economics or business explained most of the variation in ethical behavior, this empirical fact should at least trigger some reflection. Although more empirical studies are needed to compare the magnitude of the self-selection and the learning effects, what remains puzzling is the consistent association between economics and business studies and more selfish and dishonest students. These empirical results should be acknowledged and understood, not ignored or dismissed.

Third, the literature on the potentially negative effects of incentives, including their ability to crowd out intrinsic motivation and induce undesired behavior, is vast and encompasses studies by economists, organizational scholars, and social psychologists (see, for instance, Pfeffer and Sutton 2006) for a partial review). F&F’s article does not consider much, if any, of that literature. And F&F’s choice of examples in their article may not be so propitious. They cite with approval the use of “high powered” incentives described by Teece (2003) based on his experience founding and managing an expert consulting firm, LECG. But according to the firm’s 2007 annual report and 10-K, between December 31, 2006, and February 28, 2008, the stock price of LECG fell by 49%, whereas the NSDAQ composite index fell by 6% and the stock price of LECG’s peer companies increased by 46%. The decline in the value of LECG can be attributed in part to the precipitous decline in the company’s profitability, as its earnings in 2007 were about half of those in 2005 and 2006. It might be argued that the troubles at LECG stem precisely from these high-powered incentives that rewarded individual “rainmaking” but that did not reward collaboration, culture building, or institutional citizenship. It was possibly the absence of such social infrastructure that caused the firm to have problems with size and growth. Indeed, to address these problems the firm “took steps to better align expert incentives with shareholder objectives,” to quote LECG’s CEO in the letter to shareholders in the 2007 Annual Report.
Some Implications

In the end, the questions of (a) whether or not theories are self-confirming and (b) the conditions under which they are, as well as (c) the effects of theory on organizations and individuals who come to believe them, are empirical issues. In our original article (Ferraro et al. 2005), we suggested a number of ways of exploring our arguments. Recent studies have added to our knowledge of the performative nature of social theories and how to conduct empirical work on this issue. Researchers have studied, among other things, the role of game theory in the design of the Federal Communications Commission spectrum auction (Guala 2001), the construction of markets for fishing quotas in Norway (Holm 2007), and the impact of media rankings of law schools on the behavior of the schools (Espeland and Sauder 2007). There is a plethora of social and economic phenomena that might be studied using this perspective and its methods. To name just a few, the emergence of socially responsible investing practices (Gond and Palazzo 2008), the role financial theory has played in securitization and the subprime mortgage meltdown, and the effect that social networks theory and social networking technology has had on the nature of social relationships could all be productively examined using a self-fulfilling prophecy or a performativity perspective.

It is no wonder that F&F do not like the arguments we and, as we have and they have noted, many others (e.g., Ghoshal 2005, Adler 2002) have made. That is because there are some implications that if theories are self-fulfilling, academics may have different responsibilities and need to be more self-reflective about their research and scholarship.

First, if theories matter, than what we do and write and teach also matters. The thrust of the argument made by Ghoshal (2005), Khurana (2007), and others is that professional education is not just about facts, analytical methods, or even theories. Professional education is about values, and there are clearly values embedded in and implied by what we teach. Although F&F may be quite unconcerned about the fact that free riding is more prevalent among economics students than others (Marwell and Ames 1981) or that economics professors give less to charity than others (Frank et al. 1993), we are not as sanguine. Unlike our colleagues in the physical sciences, who, unless they are working on explosives, nuclear energy, contagious diseases, or genetic engineering, are unlikely to actually materially alter the world through their teaching and research, such is not the case with professors in the social and organizational sciences.

This argument implies we need to take responsibility for the consequences of what we teach and write, something already long discussed in medicine in the field of bioethics and in several of the physical sciences that do affect the environment. There could and should be more empirical studies of the effects of exposure to or belief in various theories on values and behavior, much like research on the effects of economics training. Schools could develop “ethical consequences of theory” panels, research, and expertise, paralleling what is done in medicine and in some physical sciences. Note that this is not the same as the study of business ethics and is more than the protection of human subjects; rather, it focuses on the ethical and moral consequences of what we teach and how we do our research.

However, taking responsibility, including taking responsibility for assessing the consequences of our work, is never a pleasant prospect. It is important to note that our responsibilities and our theories concern more than just the pricing of options. The brave new world unleashed by securitization might have been an important factor in the eventual financial collapse of the United States experienced in late 2008, but the role of financial innovation and its origins in finance theory was not prominent in the public debate. Or, to take another example, in a simply amazing tale of academic arrogance and its consequences, Kogut and Spicer (2005) show how dissenting views on economic reform were systematically excluded from the World Bank in its hiring and promotion practices, with disastrous consequences for Russian society and its economy. Obviously, novel social, economic, and financial practices are always a mixed bag of positive and negative consequences, but we could and should do more to think harder about the practical consequences (intended and unintended) of our work.

Second, there are implications for how we test our theories. As Henshel (1982) recognized long ago, if predictions and theories set off chains of events that can make them self-confirming, testing such theories requires more subtlety and more attention to the mechanisms that may make them appear true even if they are not. Every decent social psychologist intuitively or explicitly understands that the game is not just about figuring out what is true, it is also figuring out a set of experimental arrangements—social arrangements—that make the theory and its predictions come forth and appear.

And third, to the extent theories are self-confirming and there are multiple contending theories, multiple futures and realities are possible. To take just one simple example, the list of the most competitive economies includes not only that leader in deregulation, absence of unionization, and lack of social safety nets, the United States (ranked number 10), but also countries such as Denmark (Foss’ home country, ranked 1st), Finland (Felin’s home country, ranked 2nd), Canada (4), and Singapore (3), all of which have higher union density, more social welfare spending, and a more activist, and some might even say paternalistic, way of managing their social and economic arrangements (Economist Intelligence Unit 2008). In a similar fashion, there are many ways to build competitive companies—the contrasting
case of Wal-Mart and Costco providing just one example (e.g., Cascio 2006).

In the end, this sense of tight coupling and inevitability is what is probably most incorrect in the arguments offered by F&F. They imply strong constraints, emanating from human nature or some natural laws that cannot be altered. And by implication, they ironically imply a world with very little choice, a world in which people and organizations are constrained by the realities of one “true” social theory, independent of our theorization activity. We argue otherwise: the theories we construct do matter, and therefore, the choices we make, including the choices and ideologies embedded either implicitly or explicitly in the theories we choose to advocate, matter also. All of this means that we have the opportunity to both envision and create a different and maybe even better, more humane, and just world. But to do so, we need to be more attentive to the values and consequences embedded in the theories we embrace and more honest about their implications.

Endnotes

1 Consider as one example the “germ theory” of disease. As described in Gawande (2008), a European obstetrician in 1848 empirically demonstrated that having physicians wash their hands with a strong cleansing agent reduced the death rate among women delivering babies by an order of magnitude. But it took 31 years, until 1879, for the idea that disease could be transmitted by unseen germs and therefore preventing the spread of germs was important for health to gain widespread acceptance through the work of Louis Pasteur.

2 As we noticed in the quote beginning this article, Popper (1959) clearly distinguished between truth, which is objective and absolute, and certainty, which is subjective, and rejected the idea that we can ever know with certainty which theories are true. Indeed, as Kuhn (1970) astutely noted, Popper was not a naive falsificationist, as he writes, “In point of fact, no conclusive disproof of a theory can ever be produced; for it is always possible to say that the experimental results are not reliable, or that the discrepancies between the experimental results and the theory are only apparent and that they will disappear with the advance of our understanding” (Popper 1959, p. 50).

References


