F-Twists

Santiago Amaya

In 1953 M. Friedman published his essay, 'The methodology of positive economics'. The essay can be read as a defense of a very general methodological principle known in the literature as F-twist. In essence, F-twist can be understood as the claim that in economics, as in any other scientific discipline, theories are not to be judged by the realism of their assumptions, but by the predictive power of the hypotheses they issue.

It is not difficult to see why, since the publication of Friedman's essay, F-twist has been the center of many discussions, most of them critical of the principle. As a bold claim about scientific methodology, F-twist can be seen as simplistic and, very likely, false. First, the principle can be considered simplistic in so far as it attempts to generalize across different scientific disciplines, ignoring the peculiar methodological commitments that allegedly guide each of the very different existing research programs. Second, F-twist seems to imply the implausible claim that the sole or main virtue of scientific theories is their predictive accuracy.

F-twist, however, can also be viewed as a more circumscribed methodological principle. In order to do so, three restrictions should be observed. On the one hand, the scope of F-twist should be narrowed down to economic theories or, at least, to certain theories in the social sciences. On the other, close attention should be paid to the sort of unrealistic assumptions Friedman has in mind in the bulk of his argument. Although he

refers to assumptions concerning perfect markets, monopolies, and the like, the focus of his essay is the assumption about the rationality of economic agents. Finally, it should be noticed that even though Friedman tends to formulate F-twist in terms of predictive accuracy, he also talks about the explanatory power of economic theories, the empirical validity of their hypothesis, etc., as key theoretical virtues. Presumably, it was his idea that these virtues go hand in hand with predictive accuracy.

As such, Friedman's essay can be seen as a response to criticisms to the effect that economics is fundamentally misguided, since it assumes an unrealistic view of the nature of human beings, that is, a conception of human decision-making that is at odds with the actual psychology of human beings. These unrealistic psychological assumptions are derived from the idea that human beings are utility maximizers. Thus, under this restricted lens, F-twist can be seen as a methodological claim about economics' autonomy from psychology. Roughly, the claim amounts to saying that as long as the hypotheses of an economic theory are empirically adequate, their unrealistic assumptions concerning the rational psychology of economic agents cannot count against them. Alternatively, F-twist can be formulated as the claim that economic theories, as long as they accurately explain and predict economic phenomena, can legitimately treat human beings as if they were utility maximizers, even though they aren't actually such.

Circumscribed as it is, it is still unclear what is the precise content of the principle. The reason of this unclarity is not simply Friedman's inability to formulate F-twist in precise terms. It runs deeper than that, or so I hope to show in this paper. The thesis of

this paper is that there are several alternative ways of understating F-twist that should be carefully distinguished. These alternative interpretations of F-twist result from different ways of thinking about scientific methodology, in general, and the methodology of economics in particular. The purpose of this paper is to distinguish among these rival interpretations. It is beyond the scope of this paper to determine which is the right interpretation of the principle and, for that matter, whether it is or not a valid methodological claim. My aim is not to offer a historical reconstruction of Freidman's views, but rather to offer a map of the conceptual space, in which discussions concerning the methodology of economics and some theories within the social sciences take place.

1. F-twist$_1$

Perhaps, the best way to start our exploration is to ask in what sense the rationality assumption might be considered unrealistic. This amounts to asking what it means to treat human beings as if they were utility maximizers, even thought they aren't such. This is what Friedman says in one occasion,

> It is only a short step from these examples to the economic hypothesis that under a wide range of circumstances individual firms behave *as if* they were seeking rationally to maximize their expected returns… and had full knowledge of the data needed to succeed in this attempt; *as if*, that is, they knew the relevant cost and demand functions, calculated marginal cost and marginal revenue from all actions open to them, and push each line of action to the point at which the relevant marginal cost and marginal revenue were equal. Now, of course, businessmen do not actually literally solve the system of simultaneous equations in terms of which the

> mathematical economist finds it convenient to express this hypothesis… (1953, 223)

Friedman refers to two examples at the beginning of the quote above. The first example concerns the density distribution of leaves around a tree. According to Friedman, it might make for a good empirical hypothesis to think of the leaves *as if* they sought to maximize the amount of sunlight they receive, given their position and the position of other leaves of the same and of neighboring trees. His idea is that if the hypothesis leads to conclusions about particular trees and the position of their leaves that accord with our observations of them, then it should be a considered as an empirically valid hypothesis. The second example concerns expert billiard players, and goes along the same lines. It might make for a good empirical hypothesis, if observations coincide with the conclusions that follow from it, to think of the billiard player *as if* he knew the mathematical formulae that would give the optimum directions of travel of the balls, the angles, etc. According Friedman, there is no difference between these cases and the case of treating economic agents, *as if* they rationally sought to maximize expected utility.

Prima facie, the analogy between leaves, billiard players, and businessmen seems to be this. The hypotheses in question are hypotheses to the effect that the behavior of each of these things is going to exhibit certain patterns that can best be described as patterns in which a certain currency (sunlight, balls in the pockets, or money) is maximized. In this sense, the hypotheses are silent in relation to the mechanisms (botanical or psychological) that give rise to the behavior observed. As a consequence, the rationality assumption seems to be reduced to the idea that the behavior of economic agents has the overall

effect of maximizing expected utility. In the terminology of H. Simon (1976), the assumption seems to be reduced to a claim about substantial, rather than procedural rationality.

Now, there might be reasons for thinking that this cannot be the right way of understanding the status of the rationality assumption. These reasons are of two sorts. In the first place, it might seem questionable to treat the rationality assumption as a purely behavioral claim. To begin with, under this interpretation the notion of 'revealed preference' seems to be completely misleading, since *ex hypothesi* the assumption that agents maximize expected utility is purely behavioral. Second, and more importantly, one might doubt whether in this purely behavioral interpretation the rationality assumption is the sort of claim that can be considered realistic or unrealistic. After all, it has been argued that under suitable changes in the corresponding utility function any behavioral pattern can be described as maximizing expected utility[1].

In the second place, this behavioral reading of the rationality assumption reduces F-twist to be a very uninformative claim. In the end, the methodological claim reduces to the idea that economic hypotheses can legitimately treat agents as agents whose behavior maximizes expected utility, as long as their behavior maximizes expected utility. As a result, even if the notion of utility maximization can be spelled out in purely behavioral terms, the rationality assumption might be seen to have a very restricted scope in economic theories. To put it bluntly, if correct the assumption seems useful for yielding

[1] For a summary of the arguments presented in this paragraph, see Rosenberg 1992, pp.118-124.

predictions, but presumably insufficient for explaining particular instances of economic behavior.

This last point can be formulated in the following terms. According to the interpretation consider so far, the rationality assumption can be seen as a general principle for describing the behavior of economic agents. Assuming that this general principle is correct, it is easy to see how it yields accurate predictions. Given certain initial conditions that specify, say, conditions of the market and budgetary restrictions that affect a particular economic agent, from the principle that agents maximize expected utility, it follows that that particular agent is going to choose certain course of action –the course of action that maximizes expect utility. Logically speaking the prediction follows (deductively, perhaps) from the general principle and the initial conditions given.

The problem, however, is how to get an explanation of the behavior of the particular agent under consideration out of this. For reasons that have been adduced in discussing the nomological model of explanation associated with Hempel (1962, 1965), it seems not to be sufficient in this case to invoke the initial conditions and the general principle that economic agents maximize expected utility. In other words, for the purposes of offering an explanation it seems not to be sufficient to reverse the order of the argument that yielded the accurate prediction. First, because the resulting argument would simply describe the behavior we wanted to explain, but would not tell us why or how it happened, and these are the sort of things explanations are supposed to tell us (Kitcher 1983: 119). And second, because the explanation of a singular event requires mentioning

the causes of the event (Ruben 1990: 727), but mentioning the principle that economic agents behave rationally plus the initial conditions does not give the causes of the behavior of the agent in question.

So far, we have discussed one initial interpretation of F-twist and of the assumption of rationality, on the basis of Friedman's analogies. The upshot of this discussion is this. On the one hand, if the rationality assumption is interpreted in purely behavioral terms, then we face the challenge of blocking the claim that the rationality assumption is vacuous, since any sort of behavior can be described as rational given suitable modifications on the utility function. On the other hand, it follows from this interpretation that F-twist is reduced to the trivial claim that agents can be treated as agents whose behavior maximizes utility, only if their behavior maximizes utility. Lastly, under the interpretation proposed there are reasons for thinking that if the rationality assumption plays a major role in economic theory, then economic theories might be insufficiently equipped to provide explanations of the behavior of economic agents.

2. F-twist$_2$

Before we move on, we might want to consider an alternative interpretation of the rationality assumption also suggested by Friedman on the basis of his analogy with the billiard players. As opposed to the behavioral interpretation, this one takes the rationality assumption as claiming that agents seek to maximize expected utilities, where 'seek' has some sort of psychological import. It should be noticed how, in principle, this

interpretation might overcome the shortcomings of the behavioral interpretation discussed in the previous section. First, as a psychological interpretation it does not have to face the challenge of giving a purely behavioral description of the notions of 'revealed preference' and 'utility maximization'. Second, its psychological component makes F-twist a more robust claim about economic methodology. Lastly, the positing of a psychological mechanism of some sort possessed by economic agents gives the rationality assumption the causal aspect that might allow it to figure in the explanation of individual economic behaviors.

Here's a comment from Friedman that might take us in this direction:

> Our confidence in this hypothesis is not based on the belief that billiard players, even expert ones, can or do go thorough the process described [calculating optimum directions of travel of the balls, angles, etc]; it derives rather from the belief that, unless in some way or another they were capable of reaching the same result, they would not in fact be expert billiard players. (1953: 223)

Friedman's quote refers to the hypothesis that billiard players decide each of their shots by actually making complicated mathematical calculations. The hypothesis is a psychological one, and as such Friedman's suggest that it is false. What seems to be true, in contrast, is that whatever psychological process lead expert billiard players to shoot the way they do, the result of these process can be accounted by the hypothesis that they do make the complicated calculations.

By analogy, we can understand the use of the rationality assumption in the case of economics agents. Under this interpretation, the rationality assumption can be seen as a psychological hypothesis the truth of which is irrelevant for accounting for the behavior of economic agents, as long as the psychological mechanisms that guide them yield behavior that maximizes expected utility. Thus, in so far as this last condition is met the hypothesis could be perfectly false, but still could be legitimately used for purposes of economic theorizing. This way of understanding the rationality assumption, has two advantages that are worth noticing. On the one hand, it makes perfect sense of Friedman's claim that the rationality assumption is unrealistic. On the other hand, it makes clear in which sense economics is autonomous from psychology, and in doing so it provides Friedman's position with considerable attractiveness.

One of the challenges this interpretation faces is to explain how positing a psychological mechanism that may or may not exist might be useful for making accurate predictions of the behavior of economic agents. In this respect, Friedman hints at the possibility that there might be some force analogous to natural selection, call it 'market selection', that puts out of business those agents that do not maximize expected utility. Thus, if this force exists, whatever the psychological mechanisms that determine the behavior of economic agents are, the hypothesis that economic agents seek to maximize expected utility would always issue in accurate predictions. As it were, market selection would compensate for the unrealism of the psychological assumption. It should be observed, how this point connects with F-twist. If F-twist is taken to be the claim that economics can legitimately appeal to this false psychological hypothesis and still predict

the behavior of economic agents, an argument has to be given to the effect that 'market selection' does the work it is suppose to make[2].

Briefly, we might speculate about the form this argument might take. First, it can be an empirical argument that takes as basic premises the claim that the psychological mechanism hypothesized by economics is unrealistic but that, despite that fact, economic theories issue accurate predictions. An inference to the best explanation might suggest that what guarantees the success of economic predictions is something like market selection. Second, it can be an *a priori* argument to the effect that irrational patterns of behavior are punished by the economic system, in such a way that in the long run agents who decide irrationally disappear. Very likely, an argument of this sort would take the form of the classical Dutch-book argument originally proposed by Ramsey (1931).

It is interesting to ask whether understood in this way the rationality assumption can provide economic theories with some explanatory power. There are, at least, two reasons for being skeptical about this possibility. The first one has to do with some of the requirements frequently associated with the explanation of individual actions. As we saw at the end of the previous sections, such explanations are supposed to be causal. However, it is difficult to imagine how a psychological mechanism that might not exist can be causally efficacious of, and thus explain, the rational patterns of behavior displayed by economic agents. It seems to be a basic fact of causal relations that only real entities, systems, mechanisms, etc, can enter into them. The second reason for being

[2] As I understand it, something like this is what D. Hausman (199X) has in mind when he says that economists, as mechanics, should look under the hood.

skeptical about the explanatory power of the rationality assumption is this. If, as F-twist seems to require, there is some sort of market force that makes it the case that the psychological mechanisms actually possessed by economic agents yield maximizing behavior, then it seems that the explanation of individual rational acts should be given in terms of the market force and such psychological mechanisms. For, ultimately, these, and not the mechanism hypothesized by economic theories, are the causes of the rational behavior of economics agents[3].

In this section we have explored a second way of understanding the rationality assumption and the F-twist. The interpretation explored reads the rationality assumption as an unrealistic psychological claim and, in doing so, it provides F-twist with some special attractiveness. This interpretation faces the challenge of providing an argument for the existence of some market force that might guarantee the success of economic predictions. As we have seen, this argument might be empirical or *a priori*. Finally, we have considered the explanatory power of the rationality assumption, and have detected at least two reasons for thinking that under this interpretation economic theories might be lacking from this power.

3. Explanation and idealization

The two interpretations of the rationality assumption considered up to this point have raised doubts about the explanatory power of economic theories that rely on this assumption. This, of course, is neither a refutation of such economic theories, nor of the

[3] For variations on the arguments of this paragraph see Simon (1963).

interpretations of the rationality assumption proposed up to this point. After all, there are at least two routes the discussion above leaves open. On the one hand, it is possible to endorse everything that has been said up to this point and accept that the rationality assumption plays no explanatory role in economic theories. On the other, it is possible to endorse everything that has been said up to this point, except for the fact that explanations of human behavior ought to be causal. As far as I can see, there is nothing inconsistent with taking any of these routes.

But presumably there are also reasons for not taking any of them. Thus it might be useful to think of ways of understanding the rationality assumption so that the assumption turns out to be a legitimate expedient for the causal explanation of human behavior. The task of the sections that follow is to present some candidates interpretations of the assumption and of F-twist that meet this requirement. Before that, in this section I want to introduce some brief points on the relation between scientific idealization and explanation, which the following sections are going to exploit.

It is a fact beyond dispute that scientific theories, in general, are full of idealizations. In part, idealizations are the result of scientific attempts to present unified theories for diverse phenomena. In part, idealizations are also the result of our justified preference for simple, tractable theories. What seems to be more complicated and disputable is the right way to understand what idealizations are, and the consequences that follow from this undisputable fact about scientific theories. On the one hand, if it is conceded that idealizations are not literally true, it seems problematic to distinguish simple falsities

from idealizations. On the other, if science is full of idealizations, it is a matter of dispute whether scientific theories give us an accurate description of the world.

One common way of talking about idealization in the context of scientific methodology is to introduce the notion of 'models'. According to this view, within scientific theories, as we normally understand them, it is necessary to distinguish three different components: theories properly speaking, models, and empirical hypothesis[4]. Roughly, a theory might be understood as a conjunction of abstract axioms and postulates; empirical hypotheses are taken to be the formulations of predictions and explanations of the phenomena the theory is supposed to apply to; and models are viewed as the intermediaries that allow the transitions from the axioms and postulates of the theory to the empirical hypotheses.

The notion of 'model', however, seems to be susceptible of at least two basic interpretations (Suppes 1961). First, 'model' can be taken in a mathematical sense, a la Tarski, to refer to a set theoretic entity that satisfies all of the valid sentences of a theory. In this sense, a model is n-tuple that specifies a set of abstract objects, relations among them, and operations that range on them. Second, 'model' can be used to refer to an entity that represents the real phenomena that a theory is supposed to account for. In this sense, a model of a certain phenomenon can be a physical replica of it (as in the case of a model airplane), a picture of it (as in the case of a map or a graph), or an algebraic representation of the phenomenon. In talking about idealization in science, discussions

[4] For a very straightforward presentation of this idea, perhaps a simplistic version of it, see Forster (2000)

tend to focus on models, in the second sense defined above. We might want to refer to them, as iconic models, following F. Suppe's terminology (1973: 95ff) The common features of iconic models is that they hold with the phenomenon they are meant to model a relation of structural similarity, that is, each of the parameters that define the model can be mapped to parameters that characterize the phenomenon being modeled. Models, in this sense, are thought to be the *loci* of scientific idealization.

To understand in what sense models can be seen the *loci* of scientific idealization it is important to notice three features associated with them[5]. First, models are realizations of some of the axioms and postulates of the theories such that, literally speaking, models are true instances of them. Second, manipulations of the models consistent with the axioms and postulates of the theory aid in the discovery of theorems that follow from them, and in doing so predictive and explanatory hypothesis concerning the phenomena being modeled are generated. Lastly, even though models are structurally similar to the phenomena they are meant to represent the parameters of the model mapped to the description of the phenomena do not have to coincide. That is, this mapping relation does not have to be (and almost invariably it is not) a relation of identity but of similarity. It is because of this gap between models and phenomena that science contains idealizations. In a nutshell, idealizations are the result of the gap between the models that instantiate the theory and the phenomena they are meant to represent.

[5] In this and the next paragraph I rely heavily on Suppe (1989), especially on part II, although what I have to say is not meant to presuppose an endorsement of Suppe's Semantic conception of scientific theories.

Once the notion of scientific idealization is understood in this way, it is possible to distinguish conceptually among a variety of similarities (and dissimilarities) of different types holding among models and phenomena. These varieties might be thought to correspond to different varieties of idealizations. First, the parameters that define the model can be different from the actual parameters that characterize the phenomena, but the parameters of the phenomena might coincide (counterfactually) with those of the model (I). Second, the parameters might differ without being possible for the phenomena to coincide (counterfactually) with the model (II)[6]. Naturally, (I) and (II) correspond to different types of idealizations. Third, subdivisions might be drawn in the cases of idealizations of type (I). On the one hand, the differences between the model and the phenomena might be considered negligible (Ia). On the other, the differences between the parameters of the model and of the phenomena might be rather substantial (Ib). Fourth, in the case of idealizations of type (II) there also two possible scenarios one might think of. It might be the case that the phenomena cannot counterfactually coincide with the model, because any counterfactual adjustment in one parameter of the phenomena, relative to the model, implies a disadjustment in a different parameter, relative to the model (IIa). It might also be the case that an independent theory speaks against the possibility of the phenomena coinciding with the model (IIb). Lastly, additional subdivisions between idealizations of type (I) might be obtained by distinguishing ways in which a difference between the model and the phenomena might be considered negligible or substantial. Analogously, different ways of thinking about the counterfactual impossibility of making

[6] An example of (I) might be the assumption in Newtonian mechanics of models with frictionless plains. An example of (II) is the assumption that bodies in certain physical systems, mainly relativistic systems, are rigid. For a discussion of this second example, see Shapere (1969)

the model and the phenomena coincide might yield different subdivisions within idealizations of type (II)[7].

The important message behind these distinctions is that there are different ways of thinking about idealizations in science. More importantly, the distinction between idealizations of type (I) and type (II) provides us with the tools for making a distinction between two different ways of thinking about scientific models. We might call the models that include solely idealizations of type (I) 'approximate models', and those that also contain idealizations of type (II) 'ideal models'. Very likely, the question about whether an actual model used by a particular scientific theory is approximate or ideal cannot be decided a *priori*. The conceptual difference between these models, however, can help us draw one final distinction about different attitudes one might take towards scientific idealization.

First, one might think that only approximate models should be tolerated in scientific theories. In such case, it seems natural to think that healthy scientific practice requires detecting the differences among the parameters of the phenomena and of the model, identifying, for example, which of them are negligible and which substantial. If these differences turn out to be negligible, the predictions and explanations that the theory provides can presumably be transferred to the phenomena in a relative straightforward

[7]Frequently, both practicing scientists and philosophers of science lump together these distinctions among types of idealizations buy claiming that the application of theories to real phenomena requires *ceteris paribus* clauses. (An excellent example in this respect is Cartwright 1983.) Roughly, the idea is that predictive and explanatory hypothesis about real phenomena are derived from theories, by adding to the theorems of the theory this sort of clauses. The distinctions provided in this paragraph can be seen as an attempt to distinguish among types of *ceteris paribus* clauses.

way. If the differences are substantial, experimentation under constrained conditions –the conditions under which the putative phenomena might coincide with the model, might be used to test the empirical hypotheses that apply to the model, and these might be adjusted accordingly so as to make them applicable to the real phenomena. Experimental techniques of this sort are frequently described under the heading of the method of successive approximation (Musgrave 1981).

Second, one might think that apart from approximate models, ideal models are unavoidable in science, or at least, in the case of certain legitimate scientific theories. Idealizations of this second type, of course, cannot be treated as idealization of the former type. However, this might not mean that the theory cannot provide an account of the phenomena it is meant to explain and predict. Rather, one might think that in cases like these explanations and predictions of real phenomena can only be given in relation to the ideal model the theory provides. In such a case, the explanations and predictions concerning the phenomena might be thought to be given by reference to the explanations and the predictions that hold true of the parameters that define the model

In sum, in this section we have distinguished among different ways of thinking about idealizations in science. The account provided here is only a rough sketch of how one might go about thinking about idealizations. Rough as it is, the sketch provides with a helpful distinction between approximate models and ideal models. In the sections that follow, this distinction is put to work for the purpose of distinguishing two additional ways of thinking about F-twist and the rationality assumption in economic theory.

4. F-twist$_3$

In section 2, we considered a minimal psychological interpretation of the rationality assumption, as a way of dealing with doubts associated with the explanatory power of economic hypotheses that included this assumption. We considered the assumption as a false psychological claim, the falsity of which would not affect the empirical content of economic theories, on the proviso that certain forces of the market would compensate for it. At the end of the section, we raised further doubts about the explanatory power of the assumption, on the grounds that non-existent mechanisms cannot enter into causal explanations. With the tools gathered in the previous section we might now reformulate our interpretation of the rationality assumption to overcome this difficulty. Thus, we might want to say that the rationality assumption provides with an approximate model of the psychological mechanisms that govern the behavior of economic agents. On this interpretation, F-twist is the claim that this approximate model is enough for the purposes of constructing adequate hypothesis concerning the behavior of economic agents.

Speaking about the unrealistic assumptions included in economic theories, Friedman offers the following characterization of how economic hypotheses work:

> We can regard the hypothesis as consisting of two parts: first a conceptual world or abstract model simpler than the 'real world' and containing only the forces that the hypothesis asserts to be important; second a set of rules defining the class of phenomena for which the 'model' can be taken to be an adequate representation of the 'real world' and specifying the

> correspondence between the variables or entities in the model and observable phenomena (1953: 225)

In the case of economic hypotheses that include the rationality assumption, Friedman's description above might be interpreted in the way already suggested. The general principle that economic agents seek to maximize expected utility can be seen as summarizing an approximate model of the decision-making mechanisms of real economic agents. Under this interpretation, this principle, and the real psychological mechanism for which the model is an approximation, in addition to the correspondence rules mentioned by Friedman, should account for the empirical adequacy of economic hypotheses.

This interpretation has several advantages worth considering. First, it substitutes the claim that the rationality assumption is unrealistic, for the claim that it is an idealized model of economic decision-making, giving F-twist a less paradoxical air. Second, under this interpretation the empirical content of economic hypothesis that rely on the rationality assumption turns out to be as legitimate as the content of any scientific theory that postulates approximate models. Lastly, as long as this model can be shown to be approximate enough for the purposes of scientific theorizing, economics remains relatively autonomous from psychology.

The challenge, of course, amounts to determining whether there is some real psychological mechanism of decision-making for which the model of utility maximization can be considered an approximation. For, if the rationality assumption is

taken to describe an approximate model of the psychological mechanisms behind economic decision-making, there has to be some sort of real mechanisms that are actually being modeled. Otherwise, F-twist turns out to be simply an unwarranted claim.

In this respect, several authors (Ramsey 1931, Lewis 1976, Davidson 1992) have proposed the idea that utility maximization can be understood as an abstract version of the rationality constraints that govern commonsense psychology. In a few words, commonsense psychology amounts to the thesis that beliefs and desires are causes of human behavior, and that the relations between these are governed by constraints of instrumental rationality, mainly the constraint that people tend to choose actions which they believe to be most suited to satisfy their desires. Thus, the proposal is that the principle of utility maximization constitutes an abstract formulation of this constraint, and that utilities and probability distributions can be taken as approximate models of how desires and beliefs operate. Naturally, this proposal has not gone unchallenged. Critical discussions concerning the viability of this proposal have focused on two main points, which we can mention here only *in passim*. First, it has been debated whether utility maximization and decision theoretic utilities and probabilities are even approximate models of commonsense rationality constraints, desires, and beliefs (e.g. Pettit 1993, Hampton 199X). Second, it has been debated whether the psychological mechanisms posited by commonsense psychology have any reality after all (most notably Churchland 1981).

To summarize, the rationality assumption in economic theory can be interpreted, in the sense specified in section 3, as a shortcut for the description of an approximate model for decision-making. This interpretation might afford enough resources for providing economic hypothesis with adequate explanatory and predictive power concerning the behavior of economic agents. The challenge this interpretation faces is to ground this approximate model on real psychological mechanisms, or alternatively put, to show that this model is effectively an approximation of psychological mechanisms behind economic decision-making. In the previous paragraph a proposal in this respect has been mentioned, but it should be clear that this might be only one of the available alternatives.

5. F-twist$_4$

Finally, we arrive to one last possibility of interpreting the rationality assumption mentioned by Friedman. The backbone of this possibility is given by the notion of 'ideal model' discussed in section 3. Under this interpretation, the rationality assumption amounts to the claim that economic hypotheses might be empirically adequate even if the principle of utility maximization provides with a model of decision-making that economic agents cannot actually, or counterfactually, satisfy. It is interesting to notice that under this interpretation, F-twist amounts to the position, mentioned at the end of section 3, according to which it is legitimate for some scientific theories to gain their explanatory and predictive powers from ideal models.

Now, it makes sense to ask whether this is actually an alternative interpretation of the assumption, or whether it is just simply a version of the interpretation considered in section 3. After all, one might think that a decision-making model that is not even counterfactually realizable is not very different from a false psychological hypothesis. There is, at least, one line of reasoning might suggest that we might be in presence of a genuine new alternative. It goes as follows.

It is possible to think about ideal models as models of a normative nature. Although normative models contain more prescriptions than descriptions, normative models are not reduced to ethical considerations of the sort considered by Friedman, when distinguishing positive from normative economics. For instance, on this rather loose way of understanding the normative, logic can be considered as containing a set of normative constraints about how rational people ought to think. Analogously, one might take the rationality assumption as a normative model of how economic agents ought to make decisions. In part, it is the normative character of the model that makes it an ideal model. Now, under certain conditions models of this sort might be thought to yield some predictive accuracy. Presumably, a normative model of decision-making might provide some predictions relatively accurate, if it is the case that economic agents aim to meet the requirements of the model. Notice that this does not require the normative constraints to be reachable. It only requires the counterfactual possibility of agents being able to approximate the model up to a relative degree. On the other hand, if it is true that economic agents aim to meet the normative constraints that define the model, the model, *qua* ideal, might provide the grounds for explaining their behavior.

In a way, this line of reasoning reverses the point made by Friedman, and discussed in section 3, in relation to the plausibility of the notion of 'market selection'. As we discussed it, the rationale for appealing to market selection was to compensate for the falsity of the psychological hypothesis, in terms of which the rationality assumption was understood there. In the interpretation that is now under consideration there is no need to appeal to market selection as an economic force that takes out of business, as it were, irrational economic agents. Rather, the thought is that the gap between the normative model and the actual psychological mechanisms of real agents is closed (up to a certain extend) by their deliberate effort to approximate the ideal model.

In sum, the challenge face by the interpretation of the rationality assumption considered in this section is to make sense of the claim that ideal models are different from simple fictions. More precisely, the challenge is to show that an ideal model of rational decision-making can provide with adequate empirical hypotheses about the behavior of actual economic agents. A way of meeting this challenge has been very briefly suggested, in terms of the interplay between normative constraints and the effort real agents devote to meeting them.

6. Conclusions

The purpose of this paper was to offer different ways of understanding Friedman's twist and his claim that economic theories proceed on the basis of the unrealistic assumption

that economic agents are utility maximizers. We have considered four different interpretations, and we have reviewed some of the reasons that might support each of them and some of the challenges they face. The interpretations range from purely behavioral readings of the assumption, to psychological, and normative ones. Both the behavioral and the normative readings, perhaps the two opposite extremes, have the attractive feature that they support the autonomy of economics over psychology. The psychological interpretations discussed differ in terms of the way of understanding how literally should be taken the claim that the rationality assumption is unrealistic.