The F-Twist and the Evaluation of Political Institutions*

By Hannu Nurmi

This article focuses on the validity of an idea first expressed by Milton Friedman. According to this idea the more significant a hypothesis the more unrealistic are the assumptions underlying it. We give an overview of the recent discussion of this idea, suggest a new interpretation of it and discover some new areas of research in which its validity can be assessed.

1. The problem

During the past few decades the formal political theory has produced a number of results pertaining to the electoral process, party competition and decision making in collectivities. The ideological and metascientific views notwithstanding, one of the basic dividing lines within the community of political scientists around the world goes between those who believe that formal political theory is not only important in contributing towards an understanding of some crucial phenomena of interest, but also carries the potential of becoming *the* theory of politics in the long run, and those who suspect the relevance of the results so far achieved and, moreover, regard the whole enterprise as a futile waste of time which at best is harmless.

This paper is an attempt to evaluate the contribution of the formal political theory to our understanding of political institutions. The outcome of the evaluation does not, of course, decide the issue of superiority of one or the other of the above positions. What I hopefully can do is to elucidate the basic tenets and inference strategies of formal political theory so as to facilitate the communication between these groups.

The following specific claim commonly made by the critics of the formal political theory will be the focus of this paper: the results concerning the political institutions achieved in formal political theory are irrelevant because the actors and interaction situations analyzed in

* An earlier version of this paper was originally prepared for delivery at the Århus Joint Sessions of Workshops of the European Consortium for Political Research 29 March- 3 April, 1982 (Workshop on Regulative Political Theory). The author wishes to thank Robert E. Goodin for comments as well as the participants of the Workshop for interesting discussions.

formal theory are unrealistic. Now, as it stands this claim has been made before and, indeed, answered before in a somewhat different context. It may, therefore, be instructive to take a closer look at this previous context.

2. The F-twist

"Truly important and significant hypotheses will be found to have assumptions that are wildly inaccurate descriptive representations of reality, and, in general, the more significant the theory, the more unrealistic the assumptions (in this sense).[1]" Friedman's remarkable statement — now called the F-twist — launched a very hot debate in economics, a debate that is still continuing under various disguises. What struck the critics of Friedman was the alluded inverse relationship between significance and realism.[2]

In fairness to Friedman we must emphasize, however, that the predicates "significant" and "unrealistic" are not assigned to the same entities: significance refers to the theory, while realism pertains to assumptions. So what is at issue in the F-twist is by no means an outright contradiction in the sense that a predicate and its negation would be applied to a given entity. But anyway there seems to be something very puzzling in the statement. Before going further into the analysis of the F-twist let us remind ourselves of the fact that Friedman's position is pretty close to that of the proponents of the formal political theory. In the following I shall evaluate the criticisms directed against the F-twist from the view-point of the so-called statement view of scientific theories. Thereafter, we shall return to the formal political theory and its role in the study of political institutions. My argument runs as follows. Although the position of the proponents of the formal political theory would prima facie be easy to defend along the lines of the F-twist, I find Friedman's view simplistic and misleading. The discussion in the next four sections aims at justifying this conclusion in the light of recent debates of the F-twist. After that wa shall take a look at the evaluation of political institutions in order to point to yet another aspect of assumptions that seems to have been ignored in the debate.

3. Nagel

The inverse relationship between significance of a theory and the realism of its assumptions — the core of the F-twist — does not, however, mean that by making sure that one's assumptions are sufficiently unrealistic, one could eo ipso guarantee the significance of

[1] *Friedman* (1979), 26.
[2] See especially *Samuelson* (1963), 736.

the theory. This point is duely noticed by *Friedman*.[3] What he does not properly appreciate is the fact that there are several types of assumptions related to scientific theories. Friedman does distinguish between hypotheses and assumptions, though. The former are the building blocks of theories as the theory is in part "a body of substantive hypotheses designed to abstract essential features of complex reality".[4] What is at issue in the F-twist are the assumptions related to these hypotheses.

It seems that Friedman's "assumptions of hypotheses" are simply antecedents of conditional statements appearing either in the axioms or theorems of a theory. Or as *Nagel* puts it "in discussing Galileo's law of freely falling bodies (i.e. 'if a body falls toward the earth in a vacuum, its instantaneous acceleration is constant'), he (Friedman) asks whether this law does in fact 'assume' that bodies actually fall through a vacuum".[5] Now, is it then the case that the F-twist characterizes such assumptions? *Nagel* argues that Friedman fails to distinguish three senses in which an assumption may be unrealistic.[6] Firstly, an assumption may be unrealistic in the sense that it does not give an exhaustive description of an object or a situation. One can only wonder what kind of assumption would do that. Clearly this unrealism does not differentiate between assumptions. If this is what Friedman means by "unrealism" or "descriptive falsity", the F-twist makes sense only if one could define a measure of the distance from exhaustive description. A notion that is related to this sense of "realism" is verisimilitude.[7]

Nothing of the sort seems to have been in Friedman's mind, however. If the realism of the assumptions is understood as a dichotomous variable, we can agree with Nagel in saying that since no assumption is realistic in this sense, the F-twist is entirely uninformative.

Secondly, an assumption can be unrealistic in the sense of being false. But as Nagel observes this is unsatisfactory. Now if these kinds of assumptions are part of a theory, then obviously the theory cannot have much chance of surviving empirical testing as false assumptions entail propositions that are untrue. Such theories cannot be "significant" simply because the are rejected. If, on the other hand, the falsity characterizes the antecedent of a conditional theoretical statement, we are typically dealing either with counterfactual conditionals or with idealizations. The former types of statements have been discussed

---

[3] *Friedman* (1979), 26, fn.
[4] *Friedman* (1979), 21.
[5] *Nagel* (1979), 131.
[6] *Nagel* (1979), 133 - 135.
[7] See *Popper* (1972); *Niiniluoto* (1979).

extensively during the past decade.[8] In the social science debate the counterfactuals have not been discussed so much from the view-point of the F-twist, though. In other words, not much has been said about the general significance of theories containing counterfactuals. It is safe to agree with Friedman in saying that the presence of counterfactuals is not a sufficient condition for significance. But the issue, of course, is whether the condition is necessary. Now, in one sense it is necessary in an indirect way, viz. if the theories contain genuine law statements, they must also implicitly support counterfactuals if a specific interpretation of nomicity is adopted. That is, the very nature of a law "all x's are (necessarily) A" entails a statement "if something were an x (which it is not), it would also be A".[9] Therefore, if this view of nomicity or necessity of laws is assumed, we are committed to counterfactuals whenever we have a theory which contains law statements. But surely this does not distinguish significant from insignificant hypotheses as long as both can be expressed as law statements. To make the F-twist applicable in the present context we would have to find a way in which we could claim that the assumptions of one hypothesis are more unrealistic than the ones of another hypothesis, i.e. the entities to which a predicate is applicable are fewer than the entities to which another predicate applies. Let us see whether the F-twist would make sense in this context.

Assume that we have two hypotheses: H_1: "all entities that have the property A, have necessarily the property B", and H_2: "all entities that have property C, have necessarily the property D". Supposing that H_1 and H_2 support counterfactuals, we are committed to C_1: "if a (which is not A) would be A, then it would also be B", when entertaining H_1, and to C_2: "if b (which is not C) would be C, it would also be D", when asserting H_2.

Suppose now that the number of A-entities is less than the number of C-entities. Then clearly the assumption of H_1 is less realistic in the sense of the present discussion than that of H_2. Would it now follow that H_1 is more significant than H_2? Certainly not. All that follows is that H_2 is more generally applicable than H_1.[10]

The third sense in which an assumption may be unrealistic can be thought of as a limit of the descriptive unrealism of the type just discussed: when an assumption does not apply to anything. Nagel mentions various ideal-type constructs as examples of this kind of descriptive unrealism. We shall return to idealization shortly and

[8] See *Lewis* (1973); *Elster* (1980); *Barry* (1980).
[9] *Achinstein* (1971), 39 - 60.
[10] See also *Nagel* (1979), 134.

refrain therefore at this point from the discussion of this third type of descriptive unrealism.

Going now back to the formal political theory we notice that the rational choice models bear some resemblance to unrealism in each of the above senses: the rational actor construct is not intended to be an exhaustive description of any human being, the construct incorporates a counteractual component ("if the actors were rational in this specific sense (which they are not), then the outcome of their interaction would be the following"), and the rational actor is typically introduced as an idealization. However, when looking at the contribution the formal political theory could make to the evaluation of the political institutions, we are making normative use of the theories. We are not interested in testing the hypotheses with or without their assumptions. We are in a sense making experiments with idealized models so as to end up with predictions or outcomes that strictly speaking hold for the ideal actors and the games or other interaction settings investigated. The latter are similarly idealized. As a matter of fact, everything in models is unrealistic. And yet the very value of these experiments rests on the correspondence between models and social institutions. This correspondence justifies the inferences that are needed in the normative use of the theories.

4. Tietzel

Also *Tietzel* maintains that the F-twist is based on unsatisfactory reflection on the various types of assumptions.[11] Tietzel argues that if descriptive falsity or unrealism is understood as abstractness, there are various types of the latter and, moreover, some of these are simply unavoidable — and thus consonant with the F-twist — and some lead to an instrumentalist view of the theories.

As for the abstractness of type 1, i.e. the falsity of the antecedents in conditional statements, Tietzel maintains in accordance with what was just said in section 3 that the abstractness of this type diminishes the decidability of a hypothesis. In the case of idealized antecedents Tietzel observes that Friedman's view, which emphasizes the predictive success of hypotheses, systematically overlooks the falsifiability of the assumptions from which the predictions have been derived.[12] Hence one is led to instrumentalism. To this one might reply, however, that there is not much one could gain by testing assumptions one knows to be false. What Tietzel means is that the other premises from which — in

[11] *Tietzel* (1981).
[12] *Tietzel* (1981), 249.

conjunction with the false assumptions — the predictions are derived, are immune to empirical testing. That is, whenever the predictions from abstract (type 1) hypotheses are incompatible with observations, the modus tollens argument does not reveal anything new as we already know that at least one of the premises is false.

But in my view an even more serious situation arises when the prediction from abstract (type 1) premises is successful or as Friedman puts it, "the theory works". This means for *Friedman* that the predictions of the theory turn out to be correct.[13] Now if a theory works in this sense we are presumably not entitled to make the inference that the assumptions are false. (If we were, then we were entitled to the same conclusion regardless of whether the theory works or not.) And yet we know they are. Actually our inference possibilities can be expressed as follows:

$$\begin{array}{l} p => q \\ \sim p \\ \hline q \text{ or } \sim q \end{array}$$

That is, we could infer or predict that either q or $\sim q$ is the case. The observation q is, of course, compatible with this schema, but so also would $\sim q$ be. Indeed, whatever support q gives to the hypothesis $p => q$ or p alone, is also given by $\sim q$. It seems that as far as the unrealism in the sense of abstractness (type 1) is concerned, the F-twist is simply false.

5. Musgrave

Musgrave comes to an equally negative conclusion concerning the tenability of the F-twist.[14] While Tietzel maintains that in terms of some types of abstractness the F-twist is not strictly false but unavoidable — especially when the abstractness means the lack of exhaustive description — Musgrave regards the F-twist as untenable no matter which of his three types of assumptions is at issue.

The first type of assumptions are called the neglibility assumptions. These state that some factors which could be expected to have an effect on a phenomenon actually do not have any effect at all or at most an effect that is undetectable. As an example *Musgrave* mentions a claim sometimes made by economists in specific areas of study, viz. the assumption that there is no government.[15] Musgrave gives Friedman the credit for rightly insisting that there is no other way of evaluating a neglibility assumption than by testing the entire hypothesis system

[13] *Friedman* (1979), 30.
[14] *Musgrave* (1981).
[15] *Musgrave* (1981).

or theory. But it is one thing to argue that the realism of the neglibility assumption cannot be directly evaluated, and quite another to claim that the F-twist is true when "neglibility assumptions" are substituted for "assumptions". The latter claim is simply wrong, as Musgrave points out. The neglibility assumptions are not necessarily unrealistic or descriptively false. What they state is that some phenomena or states of affairs have a neglible effect on others. This can be quite realistic or descriptively true without affecting the significance of the theory.

Domain assumptions, on the other hand, specify the domain of applicability of the theory. Musgrave argues that sometimes neglibility assumptions are transformed into domain assumptions when the evaluation of the neglibility assumption reveals that the theory does not "work". To return to the example of the previous paragraph, it can happen that the economist finds out that the government activity has significant effects on the phenomena that his theory speaks of. He may then restrict the domain of applicability of his theory to situations where the government activity is largely absent. We observe that the change from neglibility assumptions to domain ones can go largely unnoticed even though the nature of these two types of assumptions is very different, indeed. As for the validity of the F-twist in the case of domain assumptions, we can refer back to our preceding discussion of the untenability of the F-twist because of the fact that the unrealism or descriptive falsity of the domain assumptions only guarantees restrictions on the applicability and imply nothing at all about significance.

The third type of assumptions are called heuristic assumptions by Musgrave. Their main use is in the manipulation of mathematical models when the models are very complicated. They are — if I understand Musgrave correctly — typically statements that purport to simplify e.g. a mathematical derivation of a formula by fixing certain parameters. They are also used in various thought or other experiments on models. In large computer simulation models one often performs various sensitivity analyses in order to see the effects of given parameter values on the over-all behaviour of the model. Musgrave argues that the F-twist is not valid for heuristic assumptions, either. In this, however, he does not present a clear argument. Rather the conclusion is merely stated. And yet it seems to me that with respect to the heuristic assumptions one could find some support for the F-twist. To use Friedman's example, consider the hypothesis, "under a wide range of circumstances individual firms behave as if they were seeking rationally to maximize their expected returns". Contrary to *Musgrave*[16] I would regard the part of the hypothesis starting with "as if"

as a heuristic assumption and the entire hypothesis as translatable into the following counterfactual "if the firms were seeking rationally to maximize their expected returns, then their behaviour under a wide range of circumstances would be in accordance with the observations". I don't see any change in the meaning between the two hypotheses and yet the latter one looks very much like a heuristic assumption. Moreover, whatever else Friedman wants to convey with his "as if" — clause, it seems obvious that he does not want to say that the individual firms are seeking rationally to maximize their expected returns. This is also noticed by Musgrave. So, we are back in the preceding discussion of counterfactuals. In particular, although it is difficult to decide which one of two untrue assumptions is more descriptively false, the F-twist would seem to make some sense because our example of firms suggests that a wide range of phenomena is accounted for (i.e. the theory is significant) and the assumptions are unrealistic. Of course, this does not verify the F-twist even in the case of heuristic assumptions, but I would say that the F-twist is not necessarily wrong for heuristic assumptions, whereas it is in the case of neglibility and domain assumptions.

How then does this analysis of assumptions relate to rational choice models? In other words, what types of assumptions do we encounter in formal political theory? One could argue that sometimes the rational actor is construed as a neglibility assumption. This is the case, for example, in *Riker*'s and *Ordershook*'s method of revealed preference: the analyst tries to find out the circumstances, perceptions and values of an actor that would make the observed behaviour rational.[17] The task may turn out to be impossible and the conclusion then would be that the factors assumed neglible were not neglible after all. The crux of the method is, however, that deviations from rationality are neglible. When using this method one can run into trouble and an easy way out is to modify what was considered as a neglibility assumption into a domain one. As was pointed out above, the domain assumption makes an entirely different claim: it states that certain factors certainly affect the behaviour in question, but the theory assumes that these factors either are constant or vary within a certain range. The theory does not state anything about the behaviour beyond this range. Domain assumptions have come into play in the applications of the rational choice models. It has been argued that certain domains of behaviour can be captured by means of the models because the models assume disinterested actors, while e.g. altruistic or ritualistic behaviour seems to follow from different considerations. Regardless of the validity of

[16] *Musgrave* (1981).
[17] *Riker* and *Ordershook* (1973).

this claim, we see that rationality is now regarded as a domain assumption. Incidentally, the critics of the formal political theory often seem to have this particular type of assumption in mind when the rational choice models are criticized for overly restrictive assumptions.

What about the heuristic assumptions, then? I think these abound in rational choice models. Consider, for example, the differentiable and/or separable utility function assumptions in collective goods theory.[18]

But perhaps a more relevant question is whether the rational choice model in toto can be considered as a heuristic assumption. It certainly can, as we usually construct these models in order to find out what would happen in these "unrealistic" conditions provided that the actors are rational in some precise sense. Thus, the models or the hypotheses derived from them are, indeed, unrealistic as Friedman stated. But this does not mean that the F-twist would be correct as far as the rational choice models are concerned.

As the preceding discussion shows the analysis of the F-twist hinges clearly on the meaning of "unrealistic". In particular, the F-twist seems to lose much of its intended candor if realism is considered to be a dichotomous property. The words "the more ..." in the F-twist seem to point to an idea of realism as a matter of degree. We shall therefore take a look at a notion which could possibly explicate this idea and, consequently, the F-twist.

6. Assumptions and idealizations

As far as the heuristic assumptions are concerned the unrealism or descriptive falsity of assumptions could possibly mean that the assumptions are used to "idealize" the hypotheses or laws they are linked with. The theories are then seen as consisting of statements most of which are strictly false when applied to any real research object. Moreover, the more central a hypothesis or a law in the theory, the more unrealistic it is. It seems then that the notion of idealization would fit perfectly to the F-twist. In other words, by substituting the word "idealized" for "unrealistic" in the F-twist, one could end up with a position that is entirely plausible. Of course, the latter statement depends crucially on whether one can with some accuracy characterize what is meant by idealization. We shall briefly outline an approach to idealization following *Krajewski*.[19]

[18] See e. g. *Feldman* (1980).
[19] *Krajewski* (1977); see also *Nowak* (1980), 95 - 110.

Consider an idealized hypothesis or law. It consists of two types of conditions: factual and ideal ones. These can be regarded as assumptions and denoted by A_r and A_I, respectively. Both of them are thus sets. For the sake of simplicity we consider a hypothesis or law that can be expressed as a mathematical equation

$$\forall\, x F_1\,[g_1(x), \ldots, g_n(x)] = 0$$

where $g_1, \ldots, g_n$ denote the parameters characterizing the object x and F_1 is the dependence between those parameters. For brevity we shall write the above expression as follows:

$$\forall\, x F_1(x) = 0\ .$$

If the hypothesis or law is to be idealized, it must assume some descriptively false conditions or ideal conditions, e.g. that

$$f_1(x) = 0, f_2(x) = 0, \ldots, f_k(x) = 0$$

where none of these conditions holds in any domain of application of the law. But the idealized law or hypothesis need not be entirely based on such assumptions. There may be further assumptions that are descriptively true. Hence we get the following expression for an idealized law L_1

$$L_1\colon \forall\, x\, A_r(x)\ \&\ f_1(x) = 0\ \&\ f_2(x) = 0\ \&\ \ldots\ \&\ f_k(x) = 0 => F_1(x) = 0$$

where $A_r(x)$ denotes the realistic or non-idealized assumptions about x.

Now when one tries to test a hypothesis of a similar type as L_1, one must factualize the hypothesis. This means that the idealized assumptions $f_i(x)$ $(i = 1, \ldots, k)$ are successively replaced by realistic assumptions. For example, the following sequence might ensue:

$L_2\colon A_r(x)\ \&\ f_1(x) \neq 0\ \&\ f_2(x) = 0\ \&\ \ldots\ \&\ f_k(x) = 0 => F_2(x) = 0$,

$L_3\colon A_r(x)\ \&\ f_1(x) \neq 0\ \&\ f_2(x) \neq 0\ \&\ f_3(x) = 0\ \&\ \ldots\ \&\ f_k(x) = 0 => F_3(x) = 0$

. . . . . .

$L_{k+1}\colon A_r(x)\ \&\ f_1(x) \neq 0\ \&\ f_2(x) \neq 0 \ldots\ \&\ f_k(x) \neq 0 => F_{k+1}(x) = 0$.

L_{k+1} does not contain any of the idealized assumptions of L_1. It is called the factualized law or hypothesis.

Now if the significant hypotheses are typically idealized in the above sense, then it clearly follows that the F-twist is correct as far as the unrealism of the assumptions is concerned. Indeed, the factualization process outlined above would seem to make plausible the inverse relationship between significance and realism *within a particular theory.* This is because the most significant hypotheses would seem to be the

ones mentioned in L_1. But in which sense can these assumptions be deemed most significant? I think the answer to this question shows why both Friedman and Musgrave are partly right and partly wrong.

The significance of the assumptions in L_1 is obviously related to the derivability of the hypotheses of $L_2, \ldots, L_{k+1}$ from L_1 by adding specific factual assumptions. From one point of view one could say that the process from L_1 to L_2 etc. is one of the specification of a theory because the end result is the application of the theory in a given factual context. In that sense L_1 is the most "general" of the L's. Upon closer scrutiny, however, a different picture emerges: now L_{k+1} is the most general hypothesis because all the others can be derived from it by replacing a factual assumption with an unrealistic one. Indeed, it seems to make no sense to speak of the generality of the hypotheses in $L_1, L_2, \ldots, L_k$ at all. When moving from L_i to L_{i+1} we are not "specifying" but "factualizing". So, generality in the usual sense is not decisive.

The reason why the hypothesis in L_1 is more significant than that in L_2 etc. is, according to Nowak and Krajewski, the fact that it deals with the most "essential" features of the object of study. Surely this explanation would not make Friedman very happy as he would have to reject any essentialistic ideas out of hand because of his commitment to instrumentalism. But now the F-twist would make perfect sense within a given theory. Thus willy-nilly Friedman would seem to be partly right. On the other hand, not all significant assumptions can be descriptively false. Especially if the assumptions describe the boundary conditions of the hypotheses — the domain assumptions à la Musgrave do — then they certainly cannot be false for reasons we have touched upon earlier. So, both Musgrave and Friedman seem to be partly right.

In rational choice models one could construct the idealization hierarchy of hypotheses e.g. by regarding the rational choice axioms under certainty as the idealized hypothesis L_1. The factualization would, then, start with the rational choice model under risk and continue to the uncertainty modality. Thereafter, the rational choice in strategic environments would constitute the next level and so on. Thus, $F_1(x) = 0$ consists of two axioms: 1) the 2-place relation of weak preference is complete and transitive over the set X of alternatives, and 2) for any x in X the set of alternatives inferior to x and the set of alternatives superior to x are closed sets.[20] $F_2(x) = 0$, in turn, would consist of 1) and 2') along with 3) the monotonicity in prizes axiom which states that

[20] See e. g. *Harsanyi* (1977), 31.

if A is strictly preferred to B and $p > 0$, then the lottery $(A, p; C, 1 - p)$ is strictly preferred to $(B, p; C, 1 - p)$. 2′) is the generalization of axiom 2) for risky prospects. $F_3(x) = 0$ could be expressed accordingly.

It is worth noticing that intuitively the degree of descriptive falsity diminishes when proceeding from L_1 to L_2 and from L_2 to L_3. Hence the F-twist does, indeed, apply to rational choice models. Friedman, however, fails to give any explanation of why the descriptively false assumptions are significant in those cases in which they are. I think the idealization hierarchy of rational choice theory gives a plausible explanation, viz. L_1 as constructed above is "essential" in the sense that $F_1(x) = 0$ appears in L_2 and in $F_2(x) = 0$, in L_3 and $F_3(x) = 0$ and presumably so forth, while $F_i(x) = 0$ for $i = 2, 3, \ldots$ do not in toto belong to $F_1(x) = 0$. $F_1(x) = 0$ is then the core of the system of factualized theories built "around" it. The significance of the assumptions $f_1(x) = 0 \;\&\; f_2(x) = 0 \;\&\; \ldots \;\&\; f_k(x) = 0$ lies partly in their systematic role, viz. they uncover the core or essence of the theory, i.e. $F_1(x) = 0$. But the significance is not merely due to the systematic role. L_1 seems to deal with the most obvious or unobjectionable case. The hypotheses L_2, L_3 etc. seem to deal with an extension of a particular way of structuring reality to less obvious circumstances. Indeed the plausibility of L_2 etc. would seem to lie in the plausibility of L_1.

In the preceding we have discussed the nature of unrealistic assumptions mainly from the view-point of the explanation, prediction and description of phenomena. The discussion has been pretty abstract. To cover the normative use of the theories as well as the role of unrealistic assumptions in a specific domain, we now turn to the social choice theory and its relationship to political institutions.

7. Rational actors and political institutions: examples of inference strategies

To outline the inference strategy from unrealistic assumptions in social choice theory, let us consider as an example the well-known theorem of *Gibbard* and *Satterthwaite*.[21]

The theorem says that every non-trivial, decisive and resolute social choice function with a domain of at least three alternatives is either manipulable or dictatorial. To see the inference strategy we need some standard definitions. Let X be the set of alternatives and $R = (R_1, \ldots, R_n)$ a n-tuple of individual nonstrict preference relations. The latter are assumed to be complete, transitive and reflexive. A function $F(X, R)$

[21] *Gibbard* (1973): *Satterthwaite* (1975); for a concise proof, see *Gärdenfors* (1977); see also *Feldman* (1980), 203 - 209.

which assigns to each set X of alternatives and each n-tuple R of weak preferences a nonempty subset of X, is called a social choice function. If the range of F consists of singletons only, F is called a resolute social choice function. A social choice function F is non-trivial if all elements of X belong to the range of F, i.e. for any alternative in X there is a preference n-tuple such that the alternative is chosen. We call F decisive if for all $A \subseteq X$ and for all R it yields a choice set, i.e. $F(A, R)$ is nonempty for all R and A. Let now $x P_i y$ if and only if $x R_i y$ and not: $y R_i x$. We can then define dictatorship in the customary fashion: i is a dictator iff for all $A \subseteq X$ and for all $x, y \in A : x P_i y$ implies $F(A, R) = \{x\}$. For a fixed $A \subseteq X$ and $R = (R_1, \ldots, R_n)$ we say that F is manipulable by i in the situation (A, R) iff $F(A, R') P_i F(A, R)$ where $R' = (R_1, \ldots, R_{i-1}, R_i', R_{i+1}, \ldots, R_n)$. F is manipulable (in general) iff it is manipulable by some individual in some situation.

The Gibbard-Satterthwaite theorem has the character of an impossibility theorem. It states that it is impossible to design a social choice function such that the properties of decisiveness, non-triviality, resoluteness, non-dictatorship and non-manipulability would all be present. The implication of this formulation to institutional design is: if one wishes to adopt procedures that realize a social choice function having each of these properties, one is simply wasting time. Some of the properties can be retained, but not all. But surely the theorem is based on unrealistic assumptions. It is often the case that people don't bother with ordering the alternatives so that each R_i would be complete and transitive. The "rationality" required by the theorem would thus seem to be too strong. Now, two answers can be given to this objection: (i) While it may be that this sort of rationality is not universal, it would certainly be important to cover also the cases in which all individuals are rational choosers in this sense. To put it differently, it would be strange, indeed, if the institutional designer would exclude the situations of individual rationality altogether. (ii) Both dictatorship and manipulability are defined in terms of n-tuples of R_i's, i.e. in terms of complete and transitive weak preference relations. These definitions are intuitively plausible. So much so that one could pose a counter-question: how could these notions be explicated without resort to this kind of rationality? Of course one has to bear in mind that the guidelines for institutional design are of a negative nature: the theorem states which combinations of properties of social choice function are incompatible with rationality on the part of individuals. The postulated rationality is, however, pretty weak: the actor is assumed to have a nonstrict preference order over the outcomes of alternatives.

Suppose now that one has designed a social choice procedure that is non-trivial, decisive, resolute and non-dictatorial. The Gibbard-Sat-

terthwaite theorem now tells the designers that unless special precautions are made, the procedure is in some cases unable to reveal the true preferences of the individuals provided that they are aware of the preferences of each other. If one really wants to elicit the true preferences, one can and indeed must design measures to prevent the individuals from knowing each other's preferences. Or alternatively, one could simply try to establish and/or strengthen norms that reward a truthful preference revelation.

Despite its importance as a general guideline for institutional design, the Gibbard-Satterthwaite theorem is in a way too crude a tool. To be more specific, the theorem pays no attention to the intuitive likelihood of situations which are manipulable by an individual. In other words, procedures which are manipulable under extremely special circumstances and by very few individuals are considered equivalent to procedures giving rise to strategic behaviour under almost all circumstances. In view of the fairly wide applicability of the theorem, it would be useful to have somewhat more specific information on this score. *Blair's* recent theorem is a step in this direction.[22] This theorem tells us that whenever a resolute non-dictatorial social choice function satisfies neutrality, independence and weak reduction, it is manipulable for some $A \subseteq X$ at every heterogeneous profile. Neutrality means that a relabelling of the alternatives does not affect the social choice, i.e. the same (although relabelled) alternatives are chosen after the relabelling. Independence, on the other hand, means in the present context the following. Consider a subset A of X such that $|A| = 2$ or 3. Suppose that two strict preference profiles $P = (P_1, \ldots, P_n)$ and $P' = (P_1', \ldots, P_n')$ agree for A, i.e. if we disregard all elements of $X - A$, P and P' are identical. If now these assumptions imply that $F(A, P) = F(A, P')$, F is independent. Weak reduction, finally, is satisfied by F iff $x_1 P_i x_2$ and $x_2 P_i x_3$ for all i imply that $F(X', P) = F(X'', P)$ where $X' = \{x_1, x_2\}$ and $X'' = \{x_1, x_2, x_3\}$. The theorem thus states that any resolute non-dictatorial F having these properties is manipulable for some $A \subseteq X$ whenever the strict preference profile is heterogeneous. A profile P is heterogeneous when for all possible (unordered) n-tuples P' of preferences over a set $\{a, b, c\}$ we can find a subset $\{x_i, x_j, x_k\}$ of X such that if a is substituted for x_i, b for x_j and c for x_k, the profiles P and P' agree on $\{a, b, c\}$.

Blair's theorem is interesting in relating manipulability and heterogeneity although we observe that what is given is a sufficient — not necessary — condition for manipublability for some $A \subseteq X$. Blair points out that the properties of weak reduction and independence

[22] *Blair* (1981).

actually limit the possibilities for strategic manipulation present in other types of social choice functions. The new information we get from Blair's theorem is, however, that unless the heterogeneity of preference profiles can be excluded, there is bound to be opportunities for the misrepresentation of preferences. What is therefore called for is a mechanism that excludes heterogeneity. One way of accomplishing this is to restrict the cardinality of X either by resorting to some path-independent preview process or by allowing for bargaining about the alternatives presented for social choice.

The unrealism of the above theorems is largely of the same kind. They both tell us which types of arrangements are unfeasible if one wants a sincere revelation of preferences. The descriptive falsity involved in the assumptions of these theorems is by no means of an extreme nature, intuitively speaking. The rationality required has an unobjectionable content: the rational chooser chooses that alternative he/she regards as most preferred. The only kind of behaviour one thereby excludes is the choice of an alternative that the actor knows to be less preferable than the most preferred one. It should be observed that we are now dealing with certain prospects. The unrealism of the assumptions of the theorems lies in the end only in the postulate of complete and transitive preference orders for each individual. In this respect Blair's theorem makes a stronger assumption than Gibbard's and Satterthwaite's theorem. But it could be argued that this assumption is the "essence" of rationality at least in as far as the decision and game theories are concerned, because from this notion we get more realistic or factual ones by introducing new concepts that systematize the environment of the actor.

On the other hand, the descriptive falsity of the assumption of rationality makes the result normatively binding. Firstly, if in designing a social choice procedure one must *assume* that the persons involved in the collective decision making are *irrational,* then it is certainly unlikely that the procedure will satisfy the other possibly desirable properties it is assumed to possess. This is because the persons involved would presumably learn to act in accordance with their interest at least in the long run. Secondly, if the irrationality is somehow forced upon the decision making individuals, then we are obviously dealing with an institutional design that is normatively indefensible, i.e. people are compelled to act against their interests.

8. Concluding remarks

The main conclusion from the preceding is the following: the unrealism of the assumptions can be a vice or a virtue depending on

the use of the results based on them. Musgrave is right in arguing that the F-twist cannot be true of neglibility or domain assumptions. On the other hand, it seems that Musgrave is wrong in claiming that it does not hold for heuristic assumptions, either. Of course, the descriptive falsity of the heuristic assumptions cannot alone be a sufficient condition for the significance of the hypotheses based on them, but if we consider the descriptive falsity in the sense of idealization, the F-twist would seem to make sense. Friedman would, however, probably reject the essentialistic notions involved in this rescue of the F-twist.

The formal political theory employs idealized notions, e.g. in the rational choice theory. When we use the results of rational choice theory in the evaluation of political institutions, we are not making any of the types of assumptions that Musgrave discusses. Rather we are using the rationality assumptions to make the ensuing results normatively binding. In other words, we are in effect saying that at least this type of behaviour should be taken into account as a plausible and normatively relevant type of behaviour. Consequently, the political institutions which do not work plausibly under the rationality assumptions, should be considered doubtful.

Summary

After introducing the idea called F-twist, the article deals with the recent discussion of its validity. In particular, the contributions of Nagel, Musgrave and Tietzel are focused upon. These authors argue that the F-twist simply makes too sweeping a claim and does not appreciate the various types of assumptions encountered in scientific research. We shall then outline an interpretation of the F-twist according to which the unrealism of the assumptions is viewed as idealizational in the sense of Krajewski and Nowak. This interpretation would, however, probably be unacceptable to Friedman as it would imply a rejection of instrumentalism. Finally, the paper focuses on the evaluation of social institutions by means of the results of social choice theory. It is argued that the unrealism of the assumptions plays a normative role in the evaluation. This role has largely been overlooked in the F-twist debate.

Zusammenfassung

Nach der Einleitung der Idee des F-Kniffs wird die aktuelle Diskussion über die Gültigkeit dieser Idee betrachtet. Eine besondere Berücksichtigung wird den Beiträgen von Musgrave, Nagel und Tietzel gewidmet. Diese Autoren stellen fest, daß der F-Kniff ganz einfach eine zu unbestimmte Behauptung mache und auf die Mannigfaltigkeit der Annahmen in wissenschaftlicher Arbeit ein zu kleines Gewicht lege. Danach schlagen wir eine Auffassung des F-Kniffs vor, nach der der Unrealismus der Annahmen mit dem Begriff von Idealisation à la Krajewski und Nowak verknüpft wird.

Zu dieser Auffassung würde doch Friedman wahrscheinlich eine ablehnende Haltung einnehmen, da sie mit dem Instrumentalismus im Widerspruch steht. Schließlich konzentrieren wir uns auf die Bewertung von sozialen Institutionen mit Hilfe der Theorie sozialer Wahl. Es wird behauptet, daß der Unrealismus der Annahmen eine normative Rolle in der Bewertung von Institutionen spielt. Diese Rolle hat die Diskussion über den F-Kniff größtenteils außer acht gelassen.

References

Achinstein, P. (1971), Law and Explanation. Oxford.

Arrow, K. J. (1963), Social Choice and Individual Values. 2nd ed. New York.

Barry, B. (1980), Superfox. Political Studies 28, 136 - 143.

Blair, D. H. (1981), On the Ubiquity of Strategic Voting Opportunities. International Economic Review 22, 649 - 655.

Elster, J. (1980), The Treatment of Counterfactuals. Political Studies 28, 144 - 147.

Feldman, A. M. (1980), Welfare Economics and Social Choice Theory. Boston.

Friedman, M. (1979), The Methodology of Positive Economics, in: M. Friedman, Essays in Positive Economics. Chicago 1953. (Quoted from M. Hollis and F. Hahn (eds.) (1979), Philosophy and Economic Theory, Oxford.)

Gärdenfors, P. (1977), A Concise Proof of a Theorem on Manipulation of Social Choice Functions. Public Choice 32, 137 - 142.

Gibbard, A. (1973), Manipulation of Voting Schemes. Econometrica 41, 587 - 601.

Harsanyi, J. C. (1977), Rational Behaviour and Bargaining Equilibrium in Games and Social Situations. Cambridge.

Krajewski, W. (1977), Idealization and Factualization in Science. Erkenntnis 11, 323 - 339.

Lewis, D. (1973), Counterfactuals. Oxford.

Musgrave, A. (1981), ‚Unreal Assumptions' in Economic Theory: The F-Twist Untwisted. Kyklos 34, 377 - 387.

Nagel, E. (1979), Assumptions in Economic Theory, in: A. Ryan (ed.), The Philosophy of Social Explanation, Oxford.

Niiniluoto, I. (1979), Truthlikeness in First-Order Languages, in: J. Hintikka et al. (eds.) (1979), Essays on Mathematical and Philosophical Logic. Dordrecht.

Nowak, L. (1980), The Structure of Idealization. Dordrecht.

Popper, K. R. (1972), Objective Knowledge. Oxford.

Riker, W. H. and P. C. *Ordeshook* (1973), An Introduction to Positive Political Theory. Englewood Cliffs.

Samuelson, P. A. (1964), Theory and Realism: A Reply. The American Economic Review 54, 736 - 739.

Satterthwaite, M. A. (1975), Strategy-Proofness and Arrow's Conditions. Journal of Economic Theory 10, 187 - 217.

Tietzel, M. (1981), „Annahmen" in der Wirtschaftstheorie. Zeitschrift für Wirtschafts- und Sozialwissenschaften 101, 237 - 265.