

Rational Choice, Deliberative Democracy, And Preference Transformation

Claudia Landwehr

Abstract

Within democratic theory, rational choice theories have for some time been challenged by approaches of 'deliberative democracy'. I argue that the central analytical contrast between these theories consists in their differing focus on preferences: while rational choice takes them to be stable for the time frame in question and exogenous to the decision process to be analysed, deliberative theory is concerned with their formation and transformation. The latter tends to reduce preferences to beliefs, disregarding the impact of stable and exogenous interests, while the former devotes too little attention to the formation and revision of preferences within decision processes. Where decision processes display features both of bargaining and of deliberation, a concept of preferences accommodating both stability and transformation seems to be required. I propose such a concept which regards preferences as derived and relatively stable. The concept suspends the classical distinction between desire-based and value-based preferences and incorporates both utility-maximising and argumentative rationality, claiming that we need both in every decision - the former to choose the relevant dimension and the latter to assess the options within it.

Introduction¹

Representing dominant approaches in democratic theory, rational choice theory and the theory of deliberative democracy are still widely regarded as competing approaches in more or less fundamental opposition. A common claim is that whereas rational choice constitutes an analytical and empirical theory, the idea of deliberative democracy is a purely normative one. In fact, a number of deliberative theorists have formulated their ideas in explicit opposition to the economic theories of politics which had gained influence during the

1980s. Analysts have described the development and revival of theories of deliberation since the 1980s as a reaction to the increasing dominance of rational choice theory and to its lack of normative appeal. (e.g. Bohman/Rehg, 1999; Bohman, 1998) More recently, however, a number of proposals to combine the theories has been brought forward.² The opportunity for profitable combination lies in the fact that both theories are essentially rationalist theories of action and decision. Often misperceived as a constructivist approach, the action-theoretical foundations of deliberative theory become apparent when Habermas' influential 'Theory of Communicative Action' (Habermas, 1987 [1981]) is reconsidered. Habermas' work draws heavily on speech act theory and pragmatic theory in general, which represents, as does the theory of social exchange on which rational choice theory is based, a theory of action and interaction with analytical focus on the individual.

In what follows, I will sketch what I perceive to be the central differences between the analytical assumptions of rational choice theory and those of the theory of deliberative democracy and consider in how far these differing assumptions represent conceptual and thus indispensable constituents of the respective approaches (section 2). I argue that the central contrast consists in the differing notions of preference the theories apply - rational choice theory regarding them as stable and exogenous to the decision process, deliberative democracy as transformable and endogenous. I will go on to propose an alternative concept of preferences that could combine central claims of both approaches. It conceives of preferences as derived and only relatively stable and should be regarded as a first step towards a model of preference formation and transformation in collective decision processes (sections 3.a and b). For the formation and transformation of preferences, two cognitive factors are supposed to play a role: deliberation about reasons and utility calculations. I further assume a structural similarity between preferences based on desires and preferences based on (normative) values that allows to incorporate them in the same model (section 3.c). Considering that preference-transformation through deliberation is the central idea behind deliberative democracy, I go on to present some ideas on the impact of communicative interaction on individual preferences (section 3.d). I hope that a revised concept of political preference that combines assumptions of rational choice theory and the theory of deliberative democracy may turn out to be a useful tool for analysing decision processes, especially such where bargaining and arguing seem to coexist. Some possible applications of the proposed concept of

derived and relatively stable preferences will be pointed out in the final section (section 4). I believe that the theory of deliberative democracy, which so far lacks analytical models on the basis of which operationalisable hypotheses could be formulated, might particularly benefit from a model of preference transformation, as the transformability of preferences constitutes one of its central claims. Where preference transformation indeed constitutes an empirical feature of decision-processes, such a model represents a requirement for adequate analysis and interpretation, hence indicating the need for a positive theory of deliberation to supplement the rather narrow framework of game and bargaining theory.

Central analytical assumptions of rational choice theory and deliberative democracy

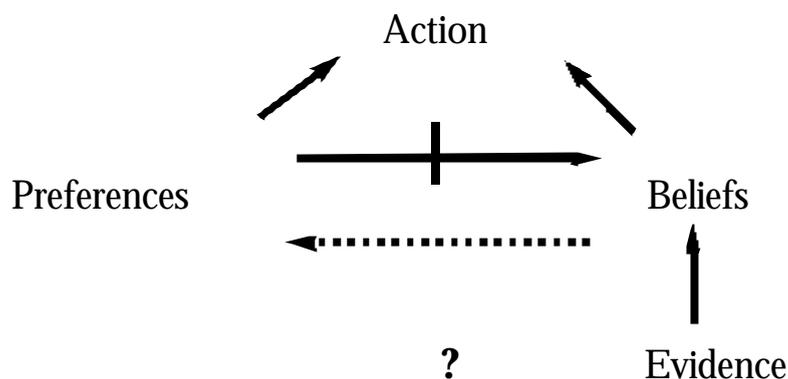
Considering the central analytical assumptions of rational choice theory and deliberative democracy, two aspects stand out as fundamental differences. The first and most important one consists in the differing concepts of rationality which the theories apply. Whereas rational choice limits its concept of rationality strictly to utility-maximising instrumental rationality, deliberative theory applies a wider concept of rationality which, in the Kantian sense, incorporates practical and theoretical reason. The latter addresses the question of what to believe while practical reason is concerned with what to do. Where rationality is conceived of as purely instrumental, the problem in decision making consists solely in finding optimal means for given ends. Hence, to be instrumentally rational, actors must only be able to predict the consequences of possible courses of action, which is in this sense a theoretical problem. Assessing the rationality of ends, by contrast, requires the wider concept of practical reason. This difference certainly has an epistemological aspect besides the analytical one to it: advocates of discourse theory in Habermas' tradition tend to claim both that individual action and decisions can be explained with reference to practical, value-oriented reason *and* that there are 'moral truths' which can be tracked in reasoned arguments in the same way as empirical facts. This epistemological difference between the two theories deserves further attention where normative claims are concerned. What is of primary interest here, however, is the analytical aspect; in order to be open to external analysis and modelling, actions need to be "accessible to rational interpretation" (Habermas, 1987, 1: 154). Empirically, arguments in which assertions are made and reasons given frequently concern norms and values as well as facts. Ethical and moral claims to validity can, as Habermas

points out, be criticised and justified in the same way as factual assertions and thus fulfill the central requirement for rationality (1987, 1: 35f). Most importantly, however, 'normative' discourses entail reasoning about individual and collective ends where the rational choice concept of rationality only allows for reasoning about means to given, undisputable ends. Whether an analytical model should make use of a wider or a more parsimonious concept of rationality obviously depends on whether or not the more limited option can provide useful generalisations and plausible interpretations in the analysis of empirical events. In the present context, it shall be sufficient to recognise the differing concepts of rationality as an important conceptual distinction between rational choice and deliberative theory.

The second important distinction is a purely analytical one and concerns the characteristics the theories assign to individual preferences. Any concept of political preference must, it seems, accommodate the twofold role of preferences pointed out by Sen (1986): preferences are both determinants of individual action and measures of individual welfare.³ Preferences may be said to relate between individual values (both normative and material) and alternative options for action in the real world. Rational choice theory explicitly defines preferences as being over possible states of the world. This, however, defines only the welfare aspect of preference. As determinants of action, preferences must be understood as preferences over alternative options, even if these are only preferred by virtue of bringing about desired outcomes.⁴ Up to this point, the theories' concepts of preference may be said to correspond to a considerable degree. Neither theory, it may be worth adding, implies assumptions on what in rational choice vocabulary would be termed the 'content of utility functions'. While rational choice is sometimes accused of misrepresenting actors as narrowly self-interested, preferences for economic benefits or social status actually represent supplementary assumptions which should not be regarded as part of a 'standard version' of rational choice theory or even conceptual constituents of the theory (cf. Brennan, 1990). Rather, the important distinction lies in the features which the theories respectively attribute to preferences on a general and more abstract level: whereas rational choice postulates preferences as stable and exogenous to the decision process itself, deliberative theory regards preferences as transformable and to a greater or lesser extent endogenous to the decision process.⁵ This difference is grounded in the differing concepts of rationality described above: where rationality is solely about means to given ends, these ends cannot be subject to a decision process. Where rational deliberation con-

cerns ends as well as means, both will be subject to transformable beliefs. Goodin defines the foundations of rational choice theory as follows: “Rational choice, as it is standardly understood, is the joint product of belief and desire. Desires dictate the ends. Beliefs merely inform us as to which means are best for serving them.” (2003: 76). Elster illustrates the same definition graphically:

Diagram One



(modified from Elster 1990: 21)

adding the requirement that preferences may not influence beliefs (which is indicated by the arrow between them being crossed out). A central question in comparing rational choice and deliberative theory, indicated by the dotted-line arrow added to Elster’s model, is whether beliefs may or should conversely influence preferences. The transformation of preferences which deliberative democracy holds possible is, after all, induced by changes in rationally grounded beliefs. Beliefs may be said to be by definition transformable, at least in so far as they entail the cognitive goal of truth or are to be instrumental for the pursuit of other goals.⁶

Leaving aside the difference in scope between the theories - deliberative theory being so far limited to democratic theory, rational choice theory being applied to a wide range of issues in numerous disciplines - as well as their differing degrees of normativity and formalisation, I maintain that any attempt

at combining rational choice and deliberative theory should have the general attributes of preferences, regardless of their content, as its prime focus. A useful starting point for this would be to consider that, rational choice being defined as the “joint product of belief and desire”, the preferences inducing this choice must similarly be a joint product of cognition and volition. This point has been made by Vanberg and Buchanan, who, with regard to constitutional choice, distinguish between an “interest” and a “theory” component of preferences:

As it is commonly *understood*, the concept of preferences is purely about *subjective values*. It refers to an actor’s evaluations of potential objects of choice. As it is commonly *used*, however, the concept typically has more than just an evaluative dimension. It is typically used in a way that blends *evaluative* and *cognitive* components, or, in other terms, that blends a person’s *evaluations* of - or *interests* in - potential outcomes of choice and his *theories* about the world, in particular his theories about what these outcomes are likely to be. (Vanberg, 1994: 168)

I will now go on to propose a concept of preferences which takes both of these components into account and systematically relates them - and which might plausibly be combined with the central assumptions of both rational choice and deliberative theory.

Towards a model of preference transformation in collective decision processes

a) Individual preferences as derived preferences

At least prior to the decision-making process, neither rational choice theory nor deliberative theory exclude any contents of preferences as ‘irrational’. A concept of preferences as preferences over alternative options for action which are preferred by virtue of the assumption that they bring about desired states of the world constitutes a useful starting point for a discussion of preference transformation. In any useful model of individual action, changes of preferences should, moreover, be restricted to changes induced by external influences or circumstances: spontaneous, random, intrinsic alterations of preferences, even if they occur empirically, cannot be made rationally comprehensible; they cannot be modelled. While a ‘thick’ concept of rationality is sometimes used to derive preferences for instrumental goods such as money or status from the rationality assumption itself, the more common

'thin' readings require empirical observation to supply the models with useful supplementary assumptions on the content of utility functions.

The most widely advocated approach for justifying such supplementary assumptions, the '*revealed* preference' approach, however, betrays an underlying idea of preferences as *derived* preferences. What an actor is thought to reveal in the actual choice of a concrete option is more than the disposition to choose a specific option under specific circumstances. The inference that under the very same circumstances (which are extremely unlikely to occur in social situations) the actor will choose the very same option, is nothing more than simple induction, which could well do without any theory of rational agency. The aim at this point is a rather more comprehensive generalisation, which, fitted into maximisation models, to enable predictions on how actors will decide in *different* situations. What such a generalisation requires is an interpretation of concrete choices as symptoms of more basic motives or desires from which action preferences are derived. Such more basic preferences are what actors are thought to 'reveal' involuntarily through their choices.

As they are not preferences over specific states of the world, but of a highly abstract nature, fundamental values and desires are pursued by choices over the respective sets of concrete alternatives. Preferences for concrete alternatives resulting in observable actions must therefore - at least implicitly - always be regarded as *derived* from concealed higher-order preferences, which may or may not be revealed in individual decisions. In fact, it is only with reference to superordinate preferences that we can plausibly speak of preferences as being ordered at all. It would seem strange and counter-intuitive to assume a preference-order in which, say, apple-pie (or a state of the world in which one has apple-pie) ranks higher than a BMW, which again ranks higher than the Conservative Party. The options are simply not comparable: "Choices must have meaning to the choosers if they are to be guided by principled considerations such as those associated with rationality." (Shepsle and Bonchek, 1997: 29) The meaning of a choice lies in the dimension with regard to which options are assessed. Only *within* a single dimension can preferences be ordered - concerning the extent to which they are instrumental to fulfil higher-order preferences from which they are derived.⁷ The relations between alternative options for action in the objective world and abstract individual values of the subjective world, however, need to be further specified. To begin with, lower-order, concrete preferences should be

regarded as derived from abstract higher-order preferences according to certain *rules*. For analytically and normatively adequate modelling, these rules, it will be argued below, require preferences to be both (relatively) stable - as Rational Choice Theory (RCT) - *and* transformable - as the Theory of Deliberative Democracy (TDD) conceives them.

b) Preferences as (only) relatively stable preferences

Considering the rules by which less abstract preferences are derived from more abstract ones, a rule of utility-maximisation appears not only plausible, but central both from a rational choice and from a deliberative perspective. If an actor has a preference for, say, a long life, both theories with their otherwise differing concepts of rationality would predict her to proceed in utility-maximising fashion in pursuing this aim. From the preference for longevity, the actor might derive one for a healthy diet, from which she derives one for food rich in vitamins, from which, finally, she derives one for fresh vegetables. Only this last derived preference can be translated into concrete action, all others remain abstract in that they still require decisions between a number of conceivable options in order to be pursued in terms of actions. The lowest-order preference resulting in action, must, it seems, be necessarily unstable. If action preferences do not change, individual action - defined as intentional behaviour - will not change and there would be no point in any theory about rationality or rational *choice*. The essential assumption of rational choice theory is, after all, that over any number of options, actors choose the one that maximises their utility. If external circumstances such as changes in prices alter the set of available options, preferences over these options will have to change as well in order for actors to rationally adopt the course of action to the new situation. Assuming that in the above case, new and reliable information that frozen vegetables contain far more vitamins than fresh ones became suddenly available, a rational actor would revise her preferences accordingly and subsequently prefer frozen vegetables over fresh ones. If she stuck to her previous preference in spite of the more utility-maximising option being available at the same cost her decision would simply be no longer utility-maximising. In situations where, as in most natural situations, information is incomplete, action preferences are always conditional preferences. The 'expected utility' of each available option and hence the preference for it depends on empirical information to be obtained from observation and communication.

But what if further reliable information on the effect of vitamins themselves became available, claiming that, contrary to priorly held beliefs, vitamins are not beneficial at all, but rather damaging to one's health? Any rational actor would then revise that more abstract preference for food rich in vitamins as well, and avoid vitamins in future. Finally, if it was discovered (and believed) that diet does not affect life expectancy altogether, rational actors simply would no longer bother about their diet. It is only the top-level 'superpreference' for a long life, which is not derived from any other preference, that remains stable regardless of changes in beliefs. Utility-maximising rationality thus implies and demands that any derived preference will be revised along with beliefs about its utility-maximising quality.

Accordingly, rational choice does not only take place between different courses of action, but also, and prior to the selection of an action, at a more abstract level. "Choice", as John Dewey put it, "signifies a capacity for deliberately changing preferences."⁸ Hence, the individual exercises her autonomy not only in actions, but also and prior to concrete action in the formation of preferences. Jürgen Habermas refers to the kind of rationality involved in selecting the right goals for concrete action as '*Wahlrationalität*' (choice rationality). Distinguishing it from instrumental rationality, he holds that these two types taken together constitute "formal rationality" – corresponding to Weber's '*Zweckrationalität*' – and contrasts it with "material (normative) rationality", corresponding to Weber's '*Wertrationalität*'.⁹ Assuming that utility-maximisation requires both instrumental and choice rationality, i.e. that it is concerned not only with utility-maximising actions but also with utility-maximising goals, *rational choice theory not only allows, but entails transformability of preferences.*

When deliberative theory describes preferences as transformable, what it means is that preferences necessarily represent beliefs as well as basic determinants of action. Some deliberative theorists might even go as far as to argue that all preferences are constituted by beliefs and therefore transformable, and might hence reduce preferences entirely to beliefs. However, such a conception would be problematic in two respects: first, preferences are, as pointed out above, not only determinants of action, but also represent standards of individual well-being (Sen, 1986: 73). If no such measure were to survive the process of deliberation, there would be no standard - exogenous to the process itself - left for an evaluation of its results. This point also has a normative aspect to it. The dangers of demagogy and manipulation,

which are highlighted by most advocates of deliberative theory, appear more where higher-order preferences are transformed to the harm of actors by such techniques. The more abstract the preference that has been changed is, the smaller would be the chance for actors to find out that they have been misled. While it tends to be easy to evaluate whether an action has served to fulfil a concrete end, it is much more difficult to estimate the effects of decisions on more abstract aims which do not immediately result in actions. Accordingly, if no higher-order preferences were to remain stable, actors would be vulnerable to systemic manipulation within a discourse without a chance of even noticing it.

A second respect in which an assumption of completely transformable preferences would be problematic pertains to the concept of individual identity. What, if not her individual preferences, constitute an actor's identity? A person's talents, cultural and personal background and relationships with others are, of course, of importance as well. But these represent restrictions and opportunities caused by more or less external factors. And while preferences are obviously influenced by such factors, a liberal perspective would require to grant them a particular degree of autonomy and authority (cf. Goodin, 2003). Claiming that preferences and beliefs are entirely shaped by the society one lives in would inevitably lead to an epistemically constructivist and analytically structuralist perspective, from which neither the epistemic advantages deliberative theorists expect from a fair discourse would be conceivable (because there would be no opportunity for epistemic division for labour) nor would the sort of actor-centred modelling advocated here make sense. If, moreover, actors themselves regard their preferences as an essential part of their identity, they could not be expected to have any motivation to enter a deliberative process if that was to transform those preferences entirely. After all, the person emerging from the process might, if preferences are regarded as constitutive for personal identity, no longer be identical with themselves. What follows from this is that regarding preferences as entirely transformable would turn the deliberative process into an end in itself - which is anything but what the theory of deliberative democracy would want it to be. *Therefore, deliberative theory not only allows, but entails an element of stability in its concept of preferences.*

What I would argue for is thus a conception of preferences as relatively stable. By relatively stable, I mean that on the one hand, derived preferences (including all preferences which could possibly be revealed) are necessarily

transformable, but that on the other hand, preferences can and should not be entirely reduced to beliefs and are therefore to be regarded as to some extent stable. The 'superpreferences' from which lower-order preferences are derived should be considered as individual dispositions for choice and action which are to some extent exogenous to processes of deliberation and collective decision-making and would tend to be particularly stable. The stability of preferences, however, should be regarded as decreasing with their level of abstraction. One reason for this is that every decision over options which more or less serve to fulfil a preference is based on beliefs which are likely to change when new information and experience become accessible. Arguments and information about the utility-maximising quality of concrete options in the objective world are easier to check up on than arguments and information concerning more abstract options, particularly where the context allows experimental action. At the same time, weighing concrete actions against one another requires complex quantitative reasoning, which is unlikely to result in certainty (Slovic, 2000). Accordingly, the revision of preferences over concrete options causes less cognitive effort than that of more abstract preferences, which are held to with a higher degree of certainty. The probability attached to preferences being 'right', i.e. to the beliefs they are based on being true, is thus likely to increase with the level of abstraction. Most importantly, however, superordinate preferences need to be stabilised in order to serve as *reasons for decisions* - even where the beliefs they are based on are problematic. Thus, with regard to the subordinate preference derived from it, a superordinate preference must constitute stable premises. This argument is supported by the fact that empirically (it seems to me), people are commonly unsure whether their actions actually serve their goals, but are much more convinced of the 'rightness' of the reasons or goals for the same action, say, the benefits from a healthy diet or a good education.

While the assumption that all preferences, including action preferences, are stable, is, as pointed out above, unsuitable for any kind of social theory, I also argue against the binary conception common in rational choice theory. This conception is one of a two-level preference structure in which 'genuine' preferences are stable while 'induced' action preferences are, being conditional on beliefs, transformable. (cf. Ferejohn, 1993; Austen-Smith, 1992) Instead, I would argue for an assumption of gradually decreasing stability of preferences. This would, for instance, make it plausible that in trying to convince somebody to alter their preferences, one would appeal to a higher-order preference, which is presumably shared or for which at least the beliefs it is based

on are agreed upon. So if I was to convince someone to buy frozen vegetables instead of fresh ones, I would appeal to the higher-order preference for food rich in vitamins in order to achieve such a revision of preferences in my interlocutor. I thereby presume a consensus on the assumption that vitamins are beneficial for one's health and should be part of a healthy diet. Both the stability and homogeneity of preferences are likely to increase with the level of abstraction and might hence raise the probability of a reasoned consensus. However, while shared higher-order abstract preferences enable consensus, dissent about fundamental values is likely to enhance conflict. The commonly stated requirement that in arguing, one should "name reasons which others can plausibly share" might nonetheless be understood to mean "name abstract preferences which others can plausibly share".

c) Desires and values

A common distinction between fundamental motives for action is that between desires on the one hand and moral values or principles on the other. Desires are commonly described as 'selfish' motives, in biological theories even equated with physical instincts, thus incidentally providing additional justification for the homogeneity assumption of universal preferences. Values, in contrast, are related to some concept of the common good or moral right, and are therefore commonly regarded as 'unselfish' or even 'altruistic' motives and sometimes judged to deserve more authority or respect than desires. In empirical contexts, however, both are described as 'interests' - e.g., trade unions and charities constituting different types of 'interest groups'. In this sense, different people with different motives have interests in states of the world which they value being realised or preserved. Hence, 'interests' could be defined as values or desires in context. While a reductionist view might try to reduce values entirely to justifications for selfish desires and while a utopian view might hope to dispel conflicting interests in favour of the common good, I would hold that it makes sense to acknowledge both but to avoid too sharp a distinction between them. As many rational choice theorists have pointed out, 'altruistic' motives can be pursued in the same utility-maximising fashion as 'selfish' ones. At the same time, it appears highly implausible to attribute all preferences to biological instincts. It is much more convincing to view them as shaped and at least partially constituted by social values, norms, and commitments. Nonetheless, both values and desires can be highly stable and so fundamental that they cannot be regarded as instrumental for other ends or preferences.

Two points may be used to illustrate this. The first is a linguistic-semantic one pointed out by Alfred Schutz. Schutz draws a distinction between two types of motives: 'in-order-to motives' and 'because motives' (1962: 69-72). He maintains that the former justify decisions instrumentally (e.g. 'I opened my umbrella *in order* not *to* become wet') while the latter serve to explain actions and decisions causally (e.g. 'He robbed the bank *because* he had a hard childhood'). Contrary to Schutz, who regards because-motives as only accessible from the point of view of an observer (including the actor looking back on and explaining his past actions), I propose to look upon because-motives as reasons for action which are subject to the actor's decision-making process. Reasons or because-motives motivate plans for action which may then be carried out instrumentally. They cannot, however, be causally sufficient for an action and they must, before and at the time the action takes place, be present in the actor's consciousness.¹⁰ Fundamental desires and values can equally constitute because-motives for action, as the following examples indicate:

- 1) "I want an ice-cream because I want an ice-cream."
- 2) "I want slavery to be abolished because it is unjust."

Although (1) is tautological it may in some cases be a suitable and maybe even the only possible justification for a preference. Neither (1) nor (2) can be interpreted instrumentally:

- 1) * He wants an ice-cream in order to...
- 2) * She wants to abolish slavery in order to...

Derived preferences, on the other hand, may be instrumental regardless of whether they are derived from desires or values:

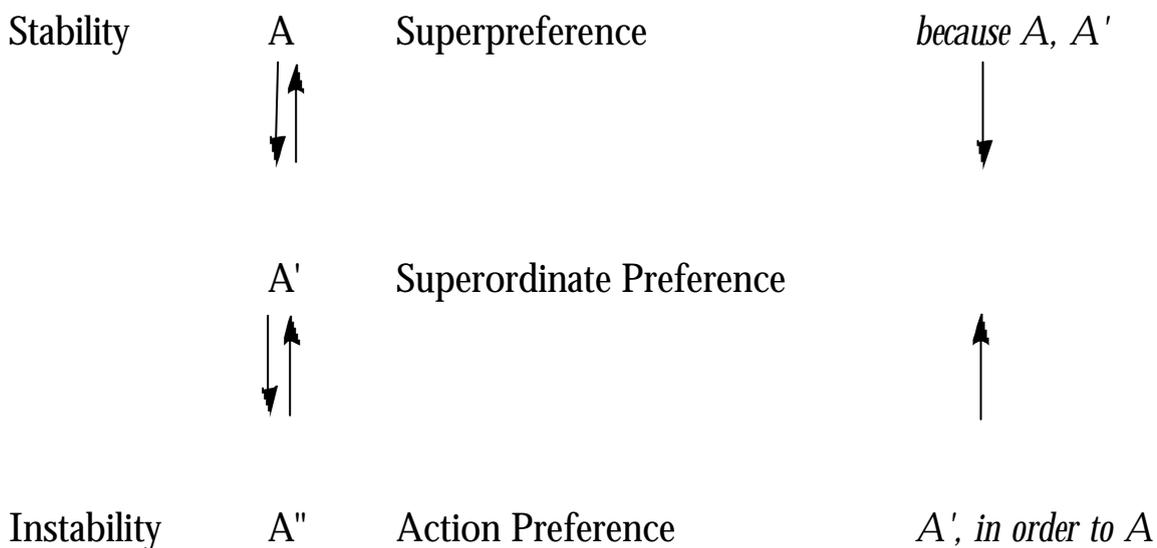
- 3) "I want a new car in order to impress my friends."
- 4) "I want education to be free in order to ensure equal opportunities."

From a consequentialist point of view (contrary to a *Gesinnungsethik* one), there even seems to be an obligation in the pursuit of normatively desirable ends to choose the instrumentally optimal options.

The second point to illustrate the similar structure and function of desires and values is an empirical one indicated by Paul Slovic (2000) and Eldar Shafir et al. (2000). They discuss how the presence of more than one reason (or superordinate preference) in a decision causes trouble in choosing options. In particular, people are apparently confused by the presence of conflicting dimensions in a decision (e.g. price and quality). It seems that contrary to rational choice beliefs, empirical actors cannot carry out a quantitative trade-off between conflicting preferences that would allow to establish an absolute utility for given options. Instead, quantitative reasoning appears possible only relative to a given dimension. Hence, actors try to establish *reasons* to guide the qualitative choice of the relevant dimension *before* they apply quantitative, utility-maximising rationality in selecting the optimal option within this dimension. Reasoning about reasons is in this sense prior to reasoning about options and their utility. A concept of preferences that regards preferences as derived and relatively stable and accounts both for the qualitative and the quantitative part of the decision process is illustrated below.

Diagram Two

Subjective World



Objective World

A bottom-up reading starting from the preference for action and moving upwards towards the superpreference it is derived from would be based on a consequential, utility-maximising logic, while a top-down reading would be based on a reason-giving, argumentative logic.¹¹ Although the former might appear more appropriate for desire-based preferences and the latter more appropriate for value-based preferences (we do something *in order to* fulfil a desire and we fight for something *because* a value we hold demands us to do so), both qualitative-argumentative and quantitative-instrumental logic are involved in most decision processes. Argumentative logic turns reasons into motives for action, hence providing them with a dual status as both cognitive and volitional. (Habermas, 1994: 188) Once reasons (and thus motives) to act have been established, instrumental logic is used to select the option that maximises utility.¹²

This decision whether and to what extent a given option actually serves a preference is, due to permanently incoming new information and to the difficulties of quantitative reasoning, always a problematic one rendering preferences over options subject to revision and therefore necessarily unstable. The reasons for choosing a dimension tend to be more stable as they are not revised on the basis of new information but on the basis of new arguments which have a less immediate impact on reasons than information has on beliefs. The greater stability of reasons compared to empirical beliefs would also explain empirical findings that considerations about opportunity costs are more likely to be ruled out where decisions are based on values: values entail stronger reasons for choice than desires do.¹³ I would insist, however, that it makes little sense to regard desire-based preferences solely in terms of utility-maximising rationality or value-based preferences solely in terms of argumentative (reason-giving) rationality - and thus to declare the former to be unstable and the latter to be stable. When a set of options becomes available, actors are likely to be unsure at first as to how to relate this set to their preferences: if, for instance, in a time of public money shortage, only either nurseries or university education can be free, a person holding a preference for equal opportunities would have to weigh the advantages of each option and would thus be applying utility-maximising rationality in a value-based decision. At the same time, higher-order preferences for, for instance, a healthy diet, tend to assume a value-like status: it is then no longer in order to live as long as possible that we eat healthily, but because we begin to attach value to a healthy diet as such. "Diet" is then treated as the relevant dimension for a number of decisions while the relevance of the dimension itself is

no longer questioned. This might relieve us from the necessity to – as epistemologically responsible persons – permanently scrutinise the coherence of our belief- and preference system, but might at the same time increase immobility and dogmatism.¹⁴ To conclude: we need and use both argumentative and instrumental rationality in every decision - the former to choose the relevant dimension and the latter to weigh the options within this dimension.

Where co-ordination and collective decisions are concerned, however, important differences between value- and desire-based preferences remain. *Shared* desires are not necessarily *joint* desires and may well constitute conflicting interests, as many of them, such as desires for status or power, can only be pursued at the cost of others. At the same time, differing desires may lead actors to endorse a decision for different reasons.¹⁵ Values, in contrast, usually entail the wish for others to share them and hence to turn them into collectively binding rules. Identical values therefore offer opportunities for cooperation and consensus, while differences between fundamental values are likely to enhance conflict.

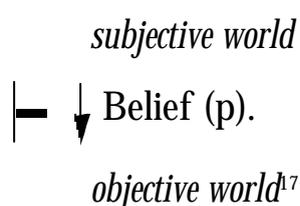
d) Discourse and preference revision

The central assumptions distinguishing deliberative from rational choice theory - its more comprehensive understanding of rationality and its claim that preferences, being based on beliefs, are transformable - are, however, not of purely analytical nature, but are based on empirical assumptions about the way discourse shapes beliefs. Rational choice theory traditionally holds that “talk is cheap” (cf. Austen-Smith, 1992), i.e. that it is no more difficult for participants in a discourse to lie than tell the truth, and consequently expects actors to do whatever serves their preferences better. Habermas, in contrast, insists that reaching agreement (*Verständigung*) is an end, a “telos”, necessarily inherent in language itself. (1987: 387; 2000: 344-347) Without denying the possibility of strategic talk such as lying, deliberative theory stresses the fact that any assertion - be it true or false, known or not known to be so - commits the speaker to certain things in certain ways.¹⁶ Where rational choice theory claims that collective decision-making is about finding solutions to collective action dilemmas and about achieving superior equilibria between competing interests, deliberative theory points out that preferences and decisions must publicly be justified with regard to universal aspects of validity - according to Habermas, truth, justice and authenticity (1987: 149). Even if they secretly pursue ‘selfish’ desires like re-election, political actors will have to

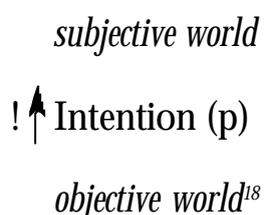
defend their positions as being pragmatically appropriate, morally right and in the common good. And even if actors only fake an interest in the common good or morally right for strategic reasons, their decisions and deeds will be measured against their words and declared preferences, not against their 'true', hidden preferences. Whereas voters could still express different preferences in a discourse than in a secret ballot, participants in bargaining rounds or deliberative commissions will eventually have to decide for what they have argued for and will be forced to yield their position if it becomes argumentatively untenable.

The observation that, as an unwritten rule of discourse, validity claims commit speakers to argumentative justification is commonplace in linguistic pragmatics and discourse theory. Being unable or refusing to provide evidence for an assertion or reasons for a request would violate the conversational 'maxim of quality' pointed out by H.P. Grice - "do not say that for which you lack adequate evidence" - and by constituting uncooperative behaviour risk a breakdown of conversation. (cf. Grice, 1975) A speaker declining to cooperate by withholding evidence and reasons simply would not be regarded as an acceptable interlocutor. This is probably the reason why even in tough bargaining situations, actors often feel compelled to justify their position argumentatively in order to prevent conversation from breaking down entirely (cf. Holzinger, 2004). At the same time, addressees are obliged to assume a positive or negative position (or an explicit abstention) with regard to any assertion or request brought by the speaker, and to defend this position in the same way the speaker defends her claim (cf. Habermas, 2000: 345).

The obligation to provide argumentative justification for requests and assertions, however, does not yet explain the influence of such arguments on beliefs and preferences. Justification could, after all, be solely a question of convention, such as greeting is. A possible explanation lies in a 'preference for true beliefs' which, as argued above, should be entailed in concepts of both instrumental and argumentative rationality. According to John Searle (1979: 12), an illocutionary act to be classified as an assertion is characterised by a "word-to-world direction of fit", which can be illustrated as:



Any assertion that is both regarded as potentially true and in conflict with currently held beliefs therefore constitutes a challenge for individual sense-making, i.e. for the assumed fit between one's thoughts (words) and the objective world. This is even more so where more complex beliefs of the type B ($p \rightarrow q$) are expressed, as arguments can to some extent be assessed with regard to their plausibility independent of the credibility of the speaker. Again taking the goal of true beliefs for granted, a plausible argument might therefore involuntarily change beliefs and preferences based on them. This is the 'force of the better argument', making it impossible to maintain beliefs in face of persuasive arguments against them. Confidence that fit between words (or thoughts) and the world is given in the prerequisite for any action intended to change the world so that it fits the idea of how one would like it to be, i.e. for any instrumental action. Rational choice models of talk tend to reduce communicative action, like all action, to the latter. The approach commonly treats all utterances as signals (cf. Austen-Smith, 1992: 46), i.e. as having the opposite, "world-to-word" direction of fit which Searle assigns to directives (including orders and claims) and commissives (including promises and threats) and which can be illustrated as:



The speaker's intention in these cases (directives and commissives) is to get the world to fit her words or, put differently, to express the wish or intention that at a point of time in the future the respectively expressed proposition (word) will fit the world.

However, even Austen-Smith in his strategic model of talk holds that "rational speakers choose their words to attempt to convince audiences to make one set of choices rather than another." (1992: 47) Yet a signal could hardly serve to 'convince' anybody. 'To convince' can only be explicated as meaning 'to get someone to hold a belief' - beliefs concerning word-to-world direction of fit. Any utterance to be used for the purpose of convincing somebody of its truth therefore constitutes, above all, a symbol. Symbols and the beliefs they represent can, however, not only refer to empirical facts but also, and often at the same time, to reasons for actions and decisions. If, as pointed out above, reasoning about ends is as much an important part of

the individual decision process as reasoning about means, a justificatory 'preference for good reasons' is likely to coincide with the instrumental 'preference for true beliefs'. Candidates for good reasons and true beliefs are, explicitly or implicitly, offered by participants in every kind of communicative interaction, including negotiations as much as deliberative arenas.

Collective decision-making is about changing states of the world. Even where such states are achieved through negotiation and compromise rather than consensus, a prerequisite will consist in a minimal set of shared and coherent beliefs, i.e. a believed fit between words or thoughts and the world. The exchange of information should therefore, involuntarily and independent of the actors' intentions, at least help to bring about a common interpretation of a given situation. As soon as actors feel compelled to name reasons for their preferences, they also offer rationales and hence motives for others to act on. Although the impact of information is probably greater than that of reasons and although persuasion by reasons will be most likely to occur in deliberative settings, it can by no means be generally ruled out in political decision-making. Being confronted with the number of options the objective world offers, actors depend both on reasons and on information in order to relate those options to their subjective world. Reasons will be needed to determine the relevant dimension (argumentative rationality) while information will be needed to estimate the instrumental value of options (utility-maximisation). I would argue that at the beginning of many decision processes, actors are unsure both with regard to the available opportunities in the objective world and with regard to the relevant dimensions for ordering them. Any teleological action, however, requires a both a degree of certainty about the objective world and its functioning and reasons to guide the decision, hence providing incentives to enter a discourse. Making assertions and giving reasons could, at least in low-cost situations, even constitute a kind of experimental action, instrumental for acquiring well-justified preferences. Reasons and information acquired in communication will then produce new and revised preferences.

Apart from the observation that in justifying their preferences, people commonly refer to the common good or morally right, it seems uncontroversial that such values play an important role as reasons for individual behaviour; what is more disputable is the question of whether it is possible, or makes sense, to include them in general models of agency (cf. Hechter, 1994). However, where having reasons for actions is an essential part of individual

decision-making and where normative values constitute particularly powerful reasons, failure to consider them in models would rob those models much of their explanatory power. Choosing preferences constitutes an essential part of individual autonomy and thus of the constitution of identity. The formation of preferences about states of the empirical and social world and the ability to justify them to oneself and to others appear to be fundamental human needs, making it reasonable for actors to enter a discourse in search both of good reasons and of reliable information. Being convinced that formerly held empirical beliefs are wrong, that a concrete option for action is instrumentally more appropriate for the attainment of a joint goal than another or that given reasons are untenable are thus processes which shape and induce revision of preferences. Reasons given by others may, moreover, induce actors to relate concrete options to different higher-order dimensions and preferences in them, e.g. to dimensions of value instead of desire or vice versa. Finally, if their own arguments or justifications are refuted, actors will be at pain to make sense of their preferences - both publicly and inwardly.¹⁹

Preference-Revision and Analysis of Political Decisions

While the claim of some rational choice theorists to be able to establish strict causal relationships between events and to make accurate predictions surely lacks both plausibility and evidence (for a criticism, see Green & Shapiro, 1994), any comparable theoretical model must similarly aim at the formulation of contestable hypotheses and interpretations. It is in this respect that rational choice theory's claim to superiority is justified and in that deliberative theory is still deficient. Whereas rational choice theory maintains that interaction may, depending on game and payoff structures, lead either to compromise on superior equilibria or persisting conflict, deliberative theory merely claims that communicative interaction bears the opportunity for reasoned consensus, implying a shift of preferences towards the pragmatically appropriate, morally right and common good. Whether or not such an epistemically superior consensus can emerge from a discourse depends on the absence of power structures and inequalities. However, to determine whether these conditions are met, the very existence of a power-free and equal discourse would be required, thus leading into an infinite regress. Some theorists who stress deliberative democracy's origin in critical theory hence regard the detection and criticism of power structures, manipulative tactics and inequalities of access and participation in discourses as the main objectives for their approach. Other advocates of deliberative democracy, howev-

er, have tried to demonstrate that their normative claims are not based on unrealistic assumptions about reality and have therefore been more concerned to show that arguing can in fact enable steps towards reasoned consensus by transforming and structuring individual preferences. (Fishkin, 1991)

Taking into account the limited time available for discussion, plausible hypotheses to be derived from a deliberative model would be that, if consensus cannot be achieved, deliberation at least reduces the overall number of dimensions deemed relevant and the number of options considered within each of them. By presenting alternative rationales for action, communication might lead actors to relate concrete options for action to dimensions of value instead of desire, allowing both for manipulation and promotion of a 'common good'. Assuming that preferences necessarily entail both reasons and beliefs, the most important consequence a deliberative model should expect from communicative interaction is a shared interpretation of the given situation. This means that even where no shared preferences for collective action can be achieved, communication about preferences, instrumentality and reasons can - from a normative perspective - still be beneficial, if it reduces the number of dimensions and options considered relevant by actors. By structuring preference orders, communication will hence at least prevent cycling majorities and enable meaningful majority decisions (Dryzek & List, 2003; Miller, 1992)

A model of derived and relatively stable preferences, that considers both argumentative and instrumental logic, might be of particular use where complex decision processes under high uncertainty, in which actors perceive a lack of information and are unsure about relevant dimensions and available options, are to be analysed. While there may be strictly deliberative forums (like expert commissions) and pure bargaining meetings (most likely non-public), many such processes are at least partly public and display instances both of strategic and of communicative action. Publicity further increases the pressure to justify one's claims and the impact of commitments entailed in utterances which are, as pointed out above, present in any kind of communicative interaction. Being unsure how different options relate to their interests and values and what the probable effects of a decision would be, political actors are unlikely to have fixed preferences over policy options. Therefore, they can be expected to revise their preferences on the basis of new arguments, new information on the opponent's preferences, expert statements, or, and probably most importantly, public opinion.

The proposed concept of preferences obviously constitutes a compromise between empirical accuracy and abstraction. While it is apparently impossible to express the influence of reasons or arguments quantitatively and to achieve a formalised representation, it is also necessary to realise an appropriate degree of abstraction. Any real actor's preference profile would certainly be far too complex for depiction: even if the number of possible super-preferences was limited, actors would derive a large number of lower-order preferences from them, which they will constantly be revising according to new information and changes in the available set. Being of analytical nature, the model can and must limit the number of relevant dimensions and preferences to be taken into consideration. In using a model of preference-transformation to explain decisions one should thus decide on a limited and fixed number of dimensions one wishes to consider. Such assumptions on relevant dimensions and actors' preferences in them are obviously supplementary assumptions which would have to be based on observation and empirical data. Subsequently, one could analyse how the arguments brought forward by single actors relate to those dimensions, what the level of abstraction of preferences they are appealing to is, how their statements commit them to certain options, and finally, how they modify others' or even their own preferences. Moreover, explicit, empirically founded assumptions about lower-order preferences would narrow down the interpretative leeway for the observer and would reduce opportunities for post-hoc modifications, which are often made where highly abstract preferences are postulated.²⁰ The plausibility of such an interpretation would obviously have to be measured against the results of bargaining and game theory, some of which are undoubtedly impressive. These approaches, however, focus on situational logic and leave little room to account for preference- and opinion shaping, which are central to democratic decision-making. Some decision processes, particularly those which are fully or partly public, in which a large number of actors are involved, which have at least some deliberative aspects besides the strategic ones and which stretch over a longer period of time, may thus not be plausibly described by those classical models. It is in such cases that I consider a model of deliberative preference-transformation useful and what I presented above are some first proposals towards such a model.

Conclusion

Complex decision processes under high uncertainty require a model of preference-shaping and preference-transformation. In such situations, it is unclear to decision-makers which alternative options exist and what reasons there are to guide the decision. It is only through new arguments and information which are gained within the decision process itself that relationships between abstract values and desires and the available set can be established. The proposed concept of preference-transformation therefore takes both aspects of preferences, the volitional and the cognitive one, equally into account. The model assumes preferences for concrete actions to be derived from fundamental desires and values in an indeterminate number of steps. As each of these steps is based on beliefs about the utility of the respective options, the degree of stability of preferences is expected to decrease with the level of abstraction. The structural similarity between value- and desire-based preferences recommends an integration of both within the same model, considering argumentative and utility-maximising logic equally for their constitution. It seems that we need and use both argumentative and instrumental rationality in every decision - the former to choose the relevant dimension, the latter to weigh the options within this dimension. The probability of preference-revision is increased by any kind of communication, as this always implies propagation of reasons and information. In fact, the constitutive rules for communicative interaction pointed out by linguistic pragmatics might support argumentative rationality and transformation of preferences. Such insights might help to specify institutional settings, interest constellations and problem types in which preference-revision may be expected to be particularly likely.

Claudia Landwehr (landwehr@sozwi.sozialwiss.uni-hamburg.de) is a PhD student and academic assistant (Wissenschaftliche Mitarbeiterin) in the department of political science at the University of Hamburg. Her doctoral dissertation deals with the transformation of actor preferences in deliberative institutions.

Notes

1. I thank Michael Th. Greven and Katharina Holzinger for helpful comments on this paper.

2. For Social Choice theory, which is commonly regarded as a subfield of Rational Choice Theory, although it does not necessarily entail its central assumptions of utility maximisation and stable preferences (see for example Miller, 1992; Knight & Johnson, 1994 and Dryzek & List, 2003). In German International Relations theory, a debate on the possibility of argumentative action in international negotiations took place in the journal *Zeitschrift für Internationale Beziehungen* in the 1990s, where a number of arguments were brought forward for a possible coexistence of strategic and communicative action, i.e. of the interactional modes of bargaining and arguing (for a survey article in English, see Risse, 2000).

3. The approach of 'revealed preference' seems to discard the link between preference and individual welfare and view preferences solely as dispositions to choose. This may make sense for consumer studies, but where political decisions are concerned, it certainly makes a difference that there is no normative point in aggregating dispositions.

4. Some might argue that the assumption of consequentialism (that options are preferred by virtue of bringing about certain outcomes) is not implied in models of deliberative democracy. A 'weak' consequentialism, though, might consider even the fact that an act has been done as a consequence of that act and hence regard even actions that constitute ends in themselves as consequential (Levi, 1991: 102). However, I understand approaches to deliberative democracy in the tradition of Rawls and Habermas as essentially consequentialist.

5. The assumption by rational choice theorists that preferences "remain fixed for the duration of the time frame in question" (Green & Shapiro, 1994: 30) is obviously of analytical rather than empirical character. However, where advocates of the approach refer to "human nature" for support of their assumptions, an instrumentally useful analytical assumption is frequently turned into a dubious empirical one.

6. Nozick argues that the evolutionary advantage in holding conscious beliefs lies in enabling human actors to adapt flexibly to changing environments (1993: 94). If beliefs were independent of evidence and experience and hence stable, they could neither be instrumentally valuable nor would they allow for a meaningful search for truth (according to Nozick, the latter possibility arose only as a by-product of the former). However, Nozick goes on to argue that where beliefs are required as premises for decisions, it will be necessary to 'take them for granted', even if always only for the *present* decision (ibid. 98).

7. Such a concept of preferences as derived preferences is proposed in Tversky's "concealed preference hypothesis" (quoted in Elster, 1997:7) and implied in Goodin's appeal to "launder preferences" (Goodin, 1986) and to "look to people's

'deeper preferences' rather than the 'superficial expressions' they give of them in everyday acts like voting" (Goodin, 2003: 49). Ferejohn (1993: 236) similarly distinguishes "genuine" preferences over "fundamental things" from "induced preferences" over "constructed entities".

8. Dewey 1937, quoted in Sunstein 1993.

9. cf. Habermas (1987: 242-247), where he discusses Weber's concepts of rationality (cf. Weber, 1964).

10. This is the view John R. Searle has recently advocated (Searle, 2001) and it appears closer to Schutz' original account of 'in-order-to' and 'because' motives in his *Der sinnhafte Aufbau der sozialen Welt* (Schutz, 1960 [1932]: 93-105).

11. March and Olsen (1989) distinguish a logic of consequentiality from a logic of appropriateness. What I mean by argumentative logic, however, differs from their logic of appropriateness in that it is rational and rationalisable. That is, validity claims to the moral rightness of a decision can be similarly made subject to contestation and justification as validity claims to the expediency of a decision. March and Olsen's logic of appropriateness rather resembles the principle behind Max Weber's 'traditional action', which is not one of rational decision.

12. Depending on the context either part of the decision may be more difficult. Searle claims that "most of the difficulty is to decide what you really want, and what you really want to do. ... [D]ecision theory ... only applies after the hard parts of the decision have already been made." (2001: 125) This is probably more true of individual decisions than of collective ones where, apart from reasons, empirical beliefs may conflict.

13. Slovic quotes the following experiment: participants are asked to choose between two road safety programmes, one of which would reduce the number of people annually killed in accidents by 100 at a cost of \$55M, the other by 30 at a cost of \$12M. An overwhelming majority choose the first option. However, in a second experiment where the price of the programme was kept secret, a majority named a price far below \$55M which they would be willing pay for it. Accordingly, they should have favoured the second option. The value-based preference to save as many lives as possible apparently featured as the relevant dimension in the decision, leading participants to neglect opportunity costs.

14. The effect seems to be similar to the "endowment effect" pointed out by Kahnemann, Knetsch and Thaler (2000): We particularly value things we currently hold (including preferences) and devalue those we do not hold.

15. On such possibilities for conflict without dissent and co-operation without consent (*Verständigung*), see Brandom (2000: 363).

16. The claim that any assertion gives rise to specific commitments on the speaker's side does not entail Habermas' assumption of agreement as an end of linguistic practice. Brandom objects to this, arguing that "[linguistic practice] makes us the kind of being we are in such a fundamental sense that it makes no sense to ask after the *point* of being like that." (2000: 363f.)

17. Modified from Searle 1979: 12.  is Frege's assertion sign.

18. Modified from Searle 1979: 14, who formalises directives as: " W(H does A)", meaning that the illocutionary act of a directive is to express that the speaker wants that the hearer (H) carry out action A and commissives as "C - I(S does A)", meaning that the speaker expresses the intention, and thereby commits herself, to doing A. I have subsumed 'want' (W) under 'intention' (I), as the authentic expression of want may be understood to imply the intention to bring about the desired state of the world (if by asking someone else to do so). Searle uses '!' for 'directive' and 'C' for 'commissive'. At least in this context, however, a commissive may be understood as a directive directed at oneself.

19. Although it is difficult to test, there are plausible arguments that actors may eventually end up having the preferences they were originally faking, possibly to reduce cognitive dissonance. Elster describes this as the 'civilizing force of hypocrisy'. (1995:250)

20. For a criticism of post-hoc modifications of assumptions in RCT, see Green & Shapiro (1994: 34-38). By the interpretative scope being too large I mean that highly abstract preferences for, e.g. wealth or power, can in many situations be used to explain entirely opposed courses of action: a 'preference for power' could well be used to explain both the resignation of a politician and their persistence in office. If it can explain either of the possible courses of action, it can explain neither.

Bibliography

- Austen-Smith, David 1992 'Strategic Models of Talk in Political Decision Making', in *International Political Science Review* Vol. 13, No. 1: 45-58.
- Bohman, James and Rehg, William 1997 'Introduction', in Bohman, J and Rehg, W (eds.) *Deliberative Democracy*, Cambridge MA, The MIT Press: ix-xxx.
- Bohman, James 1998 'Survey Article: The Coming of Age of Deliberative Democracy', in *The Journal of Political Philosophy* Vol. 6, No. 4: 400-425.
- Brandt, Robert 2000 'Facts, Norms and Normative Facts: A Reply to Habermas' [Habermas 2000], in *European Journal of Philosophy* Vol. 8, No. 3: 356-374.
- Brennan, Geoffrey 1990 'Comment: "What Might Rationality Fail to Do?"', in Cook. K.S and Levi, M (eds.) *The Limits of Rationality*, Chicago, Chicago University Press: 51-59.
- Elster, Jon 1990 'When Rationality Fails', in Cook. K.S and Levi, M (eds.) *The Limits of Rationality*, Chicago, Chicago University Press: 19-51.
- Elster, Jon 1995 'Strategic Uses of Argument', in Arrow, K et al. (eds.) *Barriers to Conflict Resolution*, New York, W.W. Norton: 236-257.
- Elster, Jon 1997 (1986) 'The Market and the Forum: Three Varieties of Political Theory', in Bohman, J and Rehg, W (eds.) *Deliberative Democracy*, Cambridge MA, The MIT Press: 3-33.
- Dewey, John 1936 'Philosophies of Freedom', in Kallen, H (ed.) *Freedom in the Modern World*, New York, Coward-McCann.
- Dryzek, John and List, Christian 2003 'Social Choice Theory and Deliberative Democracy: A Reconciliation', in *British Journal of Political Science* Vol. 33, No. 1: 1-28.
- Ferejohn, John 1993 'Must preferences be respected in a democracy?', in Copp, D; Hampton, J and Roemer, J. E (eds.) *The Idea of Democracy*, Cambridge, Cambridge University Press: 231-241.
- Fishkin, James 1991 *The Voice of the People: Public Opinion and Democracy*, New Haven, Yale University Press.
- Green, Donald P. and Shapiro, Ian 1994 *Pathologies of Rational Choice Theory*, New Haven, Yale University Press.

Grice, Herbert P. 1975 'Logic and Conversation', in Cole, P and Morgan, J.L. *Syntax and Semantics, Vol. 1, Speech Acts*, New York, Academic Press: 41-58.

Goodin, Robert E. 1986 'Laundering Preferences', in Elster, J (ed.) *Foundations of Social Choice Theory*, Cambridge, Cambridge University Press: 75-102.

Goodin, Robert E. 2003 *Reflective Democracy*, Oxford, Oxford University Press.

Habermas, Jürgen 1987 (1981) *Theorie des kommunikativen Handelns*, 2 volumes, Frankfurt, Suhrkamp (English edition: *Theory of Communicative Action*, trans. Thomas McCarthy, Boston: Beacon Press, 1984/1987).

Habermas, Jürgen 1994 (1992) *Faktizität und Geltung*, Frankfurt, Suhrkamp (English edition: *Between Facts and Norms*, trans. William Rehg, Cambridge, Polity, 1995).

Habermas, Jürgen 2000 'From Kant to Hegel: On Robert Brandom's Pragmatic Philosophy of Language', in *European Journal of Philosophy* Vol. 8, No. 3: 322-355.

Hechter, Michael 1994 'The Role of Values in Rational Choice Theory', in *Rationality and Society* Vol. 6, No. 3: 318-333.

Holzinger, Katharina 2004 'Bargaining Through Arguing: An Empirical Analysis Based on Speech Act Theory', in *Political Communication* Vol. 21: 195-222.

Kahnemann, Daniel, Knetsch, Jack L. and Thaler, Richard H. 2000 'The Endowment Effect, Loss Aversion, and Status Quo Bias', in Kahnemann, D. and Tversky, A. (eds.) *Choices, Values, and Frames*, Cambridge, Cambridge University Press: 159-179.

Knight, Jack and Johnson, James 1994 'Aggregation and Deliberation: On the possibility of democratic legitimacy', in *Political Theory* Vol. 22, No. 2: 277-296.

Levi, Isaac 1991 'Consequentialism and Sequential Choice', in Bacharach, M. and Hurley, S. (eds.) *Foundations of Decision Theory*, Oxford, Blackwell.

Lindenberg, Siegwart 1985 'An Assessment of the New Political Economy: its Potential for the Social Sciences and for Sociology in General', in *Sociological Theory* Vol. 1, No. 3: 99-113.

March, J.G and Olsen, J.P. 1989 *Rediscovering Institutions. The Organizational Basis of Politics*, New York, Free Press.

Miller, David 1992 'Deliberative Democracy and Social Choice', in *Political Studies* Vol. 40: 54-67.

Nozick, Robert 1993 *The Nature of Rationality*, Princeton, Princeton University Press.

Risse, Thomas 2000 "Let's argue!": Communicative Action in World Politics', in *International Organization* Vol. 54, No. 1: 1-39.

Schütz, Alfred 1960 (1932) *Der sinnhafte Aufbau der sozialen Welt*, Wien, J. Springer.

Schutz, Alfred 1962 *Collected Papers I. The Problem of Social Reality*, The Hague, M. Nijhoff.

Searle, John R. 1979 'A taxonomy of illocutionary acts', in *Expression and meaning Studies in the Theory of Speech Acts*, Cambridge, Cambridge University Press: 1-29.

Searle, John R. 2001 *Rationality in Action*, Cambridge MA, The MIT Press.

Sen, Amartya 1986 'Behaviour and the Concept of Preference', in Elster, J (ed.) *Rational Choice*, Oxford: Blackwell: 60-81.

Shafir, Eldar, Simonson, Itamar and Tversky, Amos 2000 'Reason-Based Choice', in Kahnemann, D. and Tversky, A. (eds.) *Choices, Values, and Frames*, Cambridge, Cambridge University Press: 597-619.

Slovic, Paul 2000 'The Construction of Preference', in Kahnemann, D. and Tversky, A. (eds.) *Choices, Values, and Frames*, Cambridge, Cambridge University Press: 489-502.

Sunstein, Cass R. 1993 'Democracy and Shifting Preferences', in Copp, D, Hampton, J. and Roemer, J.E. (eds.) *The Idea of Democracy*, Cambridge, Cambridge University Press: 196-130.

Vanberg, Victor J. 1994 *Rules and Choice in Economics*, London and New York, Routledge.

Weber, Max 1973 (1921) 'Soziologische Grundbegriffe', in Winckelmann, Johannes (ed.) *Max Weber, Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen: Mohr: 541-581.