DISCUSSION PAPERS
Institute of Economics
University of Copenhagen

03-04

Some Remarks on the Concepts of Preferences and Freedom of Choice

Hector Estrup

Studiestræde 6, DK-1455 Copenhagen K., Denmark
Tel. +45 35 32 30 82 - Fax +45 35 32 30 00
http://www.econ.ku.dk

Some remarks on the concepts
of
PREFERENCES AND FREEDOM OF CHOICE

by

Hector Estrup

INTRODUCTION

Although most people, will regard freedom of choice as a value in itself one may ask whether this love of freedom has a rational justification. For if a benign authority is able to direct a person so that his preferences are satisfied there should be no reason for him to prefer freedom of choice in favour of a submittance to the authority.. In other words why should freedom of choice be important to a person?

In a restaurant the authority, as well-aquainted with my taste, knows that I would prefer to have fried sole and that will be the dish he orders for me. What extra pleasure could I possibly have by browsing through the menu-card before the dish was ordered if, in any case, I should end up eating fried sole? I might contend that the fish- and the entire meal- would have been more becoming to my taste if I had chosen it myself: the material contents of the meal would not change but my enjoyment of it would be enhanced if I had chosen it myself. I have- so I admit- a taste for freedom.

But one may ask if such taste is reasonable, when the consequences of either my free choice or of his ordering will be exactly the same. The way a meal will satisfy my desires can be compared to the way a cure will heal my illness; and here healing is what matters and not whether the cure is prescribed to me by a doctor or it is due to some medication I have thought out on my own. So when either following orders from an authority or choosing on my own do have the same consequences there does not seem to be any rational basis for a particular taste for a free choice.

It may turn out, in some cases, that freedom of choice appears to be the better way of obtaining a fullfillment, or a satisfaction, of a person's preferences. That is, for instance, one of the ways one can derive an argument for freedom of choice when comparing a market wconomy to a command economy. In the same way Thomas Aquinas defends private propety-rights by reference to expedience.. Arguments of this type can explain a preference for freedom as an *instrument* to obtain some good, even optimal, consequences. However, they cannot give a rational foundation for a taste for freedom *as such*. And most people have a liking for freedom as a good in itself, and not only as an instrument to obtain some ulterior goods. Freedom-lovers may renounce on some other goods in order to maintain their freedom. Some are even willing to die for it. So a taste for freedom as an absolute value may sometimes be in conflict with a preference for freedom as an instrumental value. Are there rational arguments which can support a preference for freedom as such?

FREE WILL

Some may feel that there is a conflict between the idea that a person chooses freely as he prefers and the idea that we as spectators should be able to predict his choices. One may perhaps be inclined to think of a person's free will as some kind of erratic force which will make "exact" predictions of his behaviour impossible. However, there need not be any such conflict.

In economic theory a person - i.e. an agent - is characterized by his preferences, and by his abilities to perform this or that kind of work. If we know his preferences and abilities we are able to predict what he will choose to do in different circonstances. He is not forced to do as we predict. We can predict that his preferences will lead him to certain actions, but these actions will come about as a result of the way he chooses. His choices are not forced upon him by his preferences; for when someone does as he prefers,he does it willingly. What we can predict is what he *willingly* will do.

Knowing a person's preferences, his income, and prices it is possible to set up a set of demand equations which will show how the person will choose in various price-income situations.Our predictions of his choices will be based on a knowledge of these equations. In formal terms our situation is not very different from that of a scientist who can predict the movements of celestial bodies by means of a mathematical model embodying the relevant natural laws, The scientist's predictions will imply necessity. Planetary souls may or may not exist, in any case whatever "free will" such souls might have they cannot alter the celestial movements which are bound by the inviolable laws of nature. Predictions are possible because these laws are stable through time. Demand equations are stable in the same way as long as the person's preferences do not change. Preferences play the same part in the demand equations as laws of nature do in the scientist's model. And as long as a person's preferences can be assumed to stay fixed and stable it will be possible to predict how he will react, and willingly so, to changes in prices and in his income.

We may feel some uneasiness about the analogy between preferences and the laws of nature. Preferences belong to our mental life. My preferences are expressions of my will and insofar they are stable, they show that I intend to act in a systematic and consistent way. Erratic behaviour is a sign of lunacy rather than it is an expression of a free will. In a certain sense, therefore, it is the existence of a free will which makes a person's behaviour comprehensible as intentional acting and thereby also predictable.

In fact when I know a person's preferences and accordingly predict that he will choose A in favour of B I must presuppose that he has a free choice between the two and that he exerts his free will to choose A rather than B. But there is no necessity about his choice. He might choose B and so violate my prediction. If he does that there must be some facts or some features of his motivations or personality which I have not taken into account when I made my prediction: my "model" of him has turned out to be incomplete.That is also, I think, what a natural scientist would say in case of a failed experiment. It is by failed experiments, and by failed predictions, that he will learn about the lacoons in his knowledge which must be mended before he can design experiments which will succeed. Therefore when I want to predict a person's choice between different courses of actions I am in some ways in the same situation as that of a scientist, Whether my predictions will fail or succeed depends on my knowledge of the said person, i.e. knowledge of his preferences, abilities, his character, will-power and his entire situation. And when my prediction fails it may be because I do not know him, or his situation, well enough. It is insufficient knowledge, not the existence of a free will, which sometimes will make my predictions uncertain.

AGENT- OBSERVER

But a prediction of a choice may also fail for a completely different reason which has no analogy in science. If the person is aware of me he may also have an attitude regarding my endeavour to predict what he will do. His choice may be influenced by what he expects me to predict about it. If he knows my "model" of his behaviour he can predict what my prediction will contain, and then he can freely decide whether to do or not to do as he expects me to predict. So in order to predict his choice my model of his behaviour must contain his model of my behaviour, i.e. behaviour as to predicting. Is this possible without entering some sort of an infinite regres? Let us look at an example.

I know a person quite well and know that she prefers to finish a meal with a cup of coffee rather than a cup of tea. So that is what I should predict her to choose when dinner is over. But she is aware of me and hates the thought of my playing wise-guy regarding her behaviour. So in spite of her "normal" preferences she intends to ask for tea. But of course I shall take her attitude towards me into consideration when I make my prediction so I should perhaps predict her to take tea. But why should'nt she know that reasoning? And therefore eventually go for coffee. However, suspecting me to know her thoughts she should perhaps go for tea when it comes to it. Or coffee? Or tea? Or? How should it be possible for me to predict her final choice?

To an external observer it is obvious that the said person and I have entered a strategic game where I shall gain if both of us point to the same beverage whereas she will gain and I loose in all other cases. If the two of us often dine together (in spite of her animosity against my pretensions as a prophet) the game has an equillibrium in mixed strategies with each of us announcing tea or coffee with equal probabilities. Eventually my predictions of her choices will be successful half of the times we dine together. And that score is hardly a qualification for one who aspires to gain recognition as a prophet.

If the said person does not find pleasure in disappointing my prediction but on the contrary admires and reveres my fair judgment and good taste she will perhaps enhance her satisfaction by choosing in accordance with what she expects me to predict. If she expects me to predict tea she will go for tea even if she would prefer coffee had she been on her own. Sne will do as she expects me to predict, and I will make my prediction in accordance with what I expect her to expect about my prediction. The game will settle in pure strategies either coffee or tea for both of us.

In both cases, either in case of animosity or in case of admiration and reverence, the *praedicens* and the *praedictus* will enter a game with an outcome which can be analyzed by standard methods.

What this says is that a person's preferences may change when he becomes aware of being observed. But it need not affect his choosing. First, the person may be indifferent: what he intends to do he is going to do no matter what others might think or predict. And secondly, even if he is affected by being observed by other people it depends on the "strength" of his original preferences whether this will influence his behaviour. For even if it gives me great pleasure to disappoint a prophet it will not subvert my preference for honesty if he predicts that I am going to pay for the chocolate when I leave the shop. I will not steal it just for the fun of invalidating his prediction. The point, however, is that an observer cannot know in what way and to what extent his presence will influence a person's behaviour. And this ignorance is a sort of ignorance in principle because it cannot be overcome by deeper and more thoroughgoing investigations. For if I could, after laborious clinical testing, erect a formal model of his behaviour *vis a vis* me, and from which I should be able to establish predictions of his choices, this very model would, as soon as the person became aware of it, enter his universe in a way that could not, as a matter of logic, be taken care of by the very same model.

So in a way there is an infinite regres if we assume that person A has a model of person B's behaviour which include a model of person B's expectations of A' behaviour, such that the model of A's predictions corresponds the model A uses to predict B's behaviour. The regres is of the same type as the one inherent in any strategic game:

player A cannot choose an optimal strategy unless he knows how player B will choose his; and vice versa for B. In game-theory the puzzle can sometimes be solved by redefining the optimality concept, for instance by a maximin principle which makes formal analysis possible. But the regres is still there in the background because no departure from a straightforward, commonsensical concept of optimality is uncontroversial.

These problems in an agent-observer-model should be wellknown to most economists. It becomes more interesting when we realize that agent and observer can be one and the same person. This must necessarily be so when the agents are humans. The agent is assumed to have preferences, and to choose between alternatives. But as a human he must also be supposed to be able to talk about his choices, and to discuss, what he intends to do. And to do that he must act as an *external observer of his own doings*. What happens to the concept of preference, and for the possibilities for predicting his choices, if the agent is *selfconscious*?

THE SELF-CONSCIOUS AGENT

Self-consciousness comes in when we distinguish between habitual and deliberate choices. When I choose from habits I do it more or less as I am used to do and I am not conscious about any decision-making on my part. Of course when I take ccoffee instead of tea I know that this implies – at least for the moment - an abstension from tea, but this is a sheer intellectual knowledge: when I choose from habits I *know* but I do not *feel* that I choose something in favour of something else. I choose my way without feeling the dilemmas of a choice.

In a deliberate choice I am aware, sometimes painfully aware, that I am in a situation where I am forced to draw a line between what I get and what I do not get: a choice for something is *eo ipso* a choice against something else. I do not know in advance what a proper solution would be. I must think it over. weigh up *pro’s & cons´*s of the different solutions before I can find my best choice.

The usual model of consumers´ choice is well suited to an analysis of habitual choices. If we know a person well enough, his habits and his way of life, we should be able to compute how he would react if we placed him in a given universe provided it was not too different from what he was accostumed to. When the consumer chooses between coffee and tea he is not a mindless habit-automaton; he may be *conscious* about his choice, but he need not be *self-conscious* about it in the sense that he can talk about it and make arguments in favour of his choice. A normal consumer is able to choose in a sensible way, and to be conscious about it, without knowing anything about preferences, indifference-maps, budget-lines and optimation. And noone could possibly contend that a normal consumer unknowingly had all these things in his head. Take as an example a cat which jumps in order to catch a mouse. It must be able to control the trajectory of its moving body so that its paws will land exactly on the top of the mouse. This requires the solution of some very complicated differential equations which the cat knows nothing about even though it succeeds most of the times in catching the mouse. Of course it is possible to say that the cat, when it jumps, is aware of the mouse and conscious of the delects of a mouse-catch, but it is unable to analyze and discuss the way its jump is performed. So it cannot be self-conscious about it.

In a deliberate choice things do not come that easy. Different courses of actions must be analyzed, consequences computed and compared, hidden preferences must be dug up, and finally a decision must be taken which is, all things considered, the best which can be done. What I am after, deliberating my choice, is a model of my own behaviour whicn can tell me what to do.

Now suppose I succeed in setting up such a model. Then this model will be part of my deliberations as to what I shall choose. I will follow the model and decide accordingly if I find it convincing that the best decision, as indicated by the model, is the best thing for me to do. But then my decision to do as the model tells me cannot be described by that model. Any model of a deliberate decision which is known to a decisor may or may not convince him. If he is convinced the model will count among the reasons he can refer to when and if he is going to justify what he has decided to do. Therefore what he eventually decides cannot be computed from a (formal) model because any model whatsoever which pretends to describe his choice and of which the decisor is aware, can at most serve as an argument in favour of his decision. Perhaps we can say that any (formal) model of how a self-conscious agent will make his choice, or take a decision, must be *incomplete*.

In the elementary theory of consumers' choice the consumer's preferences are taken as given. Is this model "incomplete" if the consumer is self-conscious?

Self-consciousness implies that the consumer knows the model, in particular what preferences he is supposed to have. As mentionned above preferences belong to our mental life. They are not natural phenomena like height and weight. No medical or biological examination of my entrails will reveal any preferences, and I do not refer to a psycho-medical report in order to convince myself and others of the "truth" of my preferences or of my taste. Preferences are mental phenomena and are as such, at least to a certain extent, subjects to my will.

When I consider my habits, my taste and my preferences I may not like what I see. When I scrutiny my way of life I shall perhaps see nothing but bad habits, obscene taste and criminal preferences. And who says there is nothing to do about it? *De gustibus non est disputandum* is a dictum too defaitist when the object is my own tastes and propensities. It should be possible for a strong character to mobilize sufficient will-power to change his way of life. Any time time a person is aware of his preferences it becomes a problem for him if he shall live his life and do his choosing and deciding according to these preferences or if he, by character and will-power, should try to change them, to subdue his bad habits and supress lugubre and obscene interests. For a free choice will imply that man is responsible for the decisions he takes and then he will also be responsible for accepting the preferences which have led him to his decisions. The egocentric response ”That´s typically me!” can never be invoked as either an *excuse* or as a *justification* for a bad decision, although it may sometimes serve as an *explanation*.

What we are hinting at is that there is *an open end* in the preference concept when the agent is self-conscious. I may prefer A (chocolate) to B (fruit). Call this preference P . It might have been the other way round, that I preferred B to A. Call this might-have-been preference Q . If I am aware of the fact that I am under influence of P it become an open question whether I would prefer to be influenced by P or Q in my decision-making. In other words do I prefer to prefer A to B in favour of preferring B to A? This preference for preferences is a second-order preference, but nothing will prevent us from going on with preferences of still higher orders, and there is- in principle- no end to the proces. It is in this way that any decision-model for a person, *in casu* a consumer, which takes preferences as given will be incomplete if the person is self-conscious about his decision-making. Preferences cannot be taken as given if the agent is self-conscious. Then the assumption that preferences are stable may not be tenable.

CONSIDERED PREFERENCES

This problem of self-conscousness may be circumvented by what has been called *considered* preferences. Often our experience of preferences will present itself to our minds in a rather confusing manner: sometimes I prefer chocolate to fruit, then fruit to beer, but then, after all I would be inclined to prefer beer to chocolate. If I just open my mind as a thoughtless receiver to my feelings of preference for this and for that I shall find myself in a wilderness of uncoordinated binary preference-relations. Just as when you present a dog with three dishes of food in three different places in the room: the dog will run from dish to dish, still preferring one for the other, yet not knowing which is best.

When I choose coffee in favour of tea I must be aware of the alternatives. But I need not be aware of my preference for coffee over tea. However, by repeated experience of such choices I may, by reflection, be aware of it in the sense that I can talk about it to myself and to others. And in talking about my preferences I am submitted to the conceptual logic of expressions like "I prefer A to B" as it appears in ordinary language. Otherwise I shall not be able to make myself understood. This means that I am bound to respect rules of antisymmetri and transivity. If I said that I preferred A to B, B to C and C to A people would be confused and perhaps conclude that I used "prefer" in an either misleading or incomprehensible way. Of course I may prefer A to B in some situations, B to C and C to A in other situations, but that is no violation of transivity just as there is no violation of antisymmetri when I state that I sometimes prefer coffee to tea, sometimes tea to coffee. Nor will the phenomena of regret invalidate the logic of preference because the actual choice I make may be felt as a change of my situation: It is true that I prefer coffee to tea, but having got the coffee I might think it would have been nicer with a cup of tea instead and that was what I should have chosen. Here is no violation of the logic of choice because that logic, i.e.the logical grammar of the word "prefer", refers to a choice in a given situation. But when I have in fact chosen coffee -in accordance with my preferences before my choice was made my actual choice has changed my situation, and perhaps I shall become disappointed when I experience the situation I have brought myself into. But even if there was a poor soul who never failed to regret what he had just chosen it would be misleading to talk of inconsistency. There need not be anything wrong with his preference-logic. What fails him is character and will-power, that is, in a way, he lacks the ability to make responsible choices. For he will run into difficulties any time he tries to explain his decisions, i.e. to explain to others why he did not do what he now says he should have done.

Now, as a self-conscious chooser I build up, by experience, reflection and talk with others, an idea of my preferences. I see to it that I can speak in a consistent way about them. My preferences do not come to me by some sort of introspective revelation. It is from experience that I learn about my inner longings, my passions and appetites, and consulting reason I try to organize my answers to the question "what do I prefer in this or that situation?" in a pattern consistent with the logical grammar of the verb "to prefer". In the same proces I try to apply wisdom and judgment so as to subdue, if not eradicate, too much influence from what I may find in my inner self of wild and wicked passions and appetites. Eventually I will end up with a set of preferences, a utility function perhaps, which will guide my choosing, and which I can present to myself, and to others as well, without any feelings of shame and remorse. In this learning proces the combined forces of passion, reason and will have led to a structure of preferences which one may call *considered* preferences.

FREE WILL AND CAUSATION

Still there is an incompleteness as long as we do not know why a person´s considered preferences are the preferences he accepts and wants to live up to in his choosing. For considered preferences do not constitute a system of unconscious propensities but are the preferences the agent has chosen "to have". And this choice cannot be explained, or justified by a reference to the very same system of preferences. So the above-mentionned *open end* in the preference concept is still there even when the agent directs his choosing according to his considered preferences.

On the other hand it is this open end which makes it possible to consider the agent as an individual with a *free will*. For if all of the choices an agent could make could be explained by a suitable model, and if we believed in that model, then we would be bound, as some psychologists and biologists perhaps are, to consider the agent as some sort of a psycho-biological machine, or as a mindless robot. And this would be self-defeating; for the model should explain the agent's decisions, but a robot does not take decisions.the way humans do: a robot *reacts*, a person *decides*. To a selfconscious person there will always be an *a-rational* gap between a conclusion derived from deliberations as to what he should do on the one hand and on the other what he eventually decides to do. That gap cannot be bridged by rational means, i.e. by a model, for any such model could at most serve as an element in the deliberations before he made his choice. By a model we may *predict* his behaviour, but we cannot *determine* his decisions, for he is not forced to do as the model tells him to do.

Perhaps we can say, that the term ”free will” refers, *not to some particular kind of causation*, but quite simply to the existence of that gap. Then we could also say that ”free will” is a term we apply in ordinary language as a name for the void which must exist between deliberations and decisions, and which is relevant as far as humans are concerned, but not so for robots.

A self-conscious agent will see his considered preferences as the proper foundation for his choosing and for the way he will act and behave. Such agent does not see himself in a Humean way as a slave or a helpless prey to passions and inborn instincts. When he lives in accordance with his considered preferences it is because he *wills* it. They represent, or express, the way the he wants to live his life in the format of *rules* which he accepts and which, when followed, will allow him to live the good life as he sees it. They show how he ”wills” things to be if he had the choice. Of course, lapses of character, slips of memory and fits of unreason if not sheer thoughtlessness may occasionally disturb his endeavour to live up to his considered preferences. But if we know him as a reasonably strong-willed character we should in many cases be able to predict his behaviour, given that we know these preferences

PREFERENCES ARE DEFEASIBLE

Now consider a person with well-considered preferences. By forming these preferences he has freely decided how he intends to choose, act and behave in different situations. Therefore, if he is allowed to live in accordance with these preferences he cannot demand more freedom in his life because his love of freedom has already been expressed in these preferences. They represent, so to speak an embodyment of his will. What then could such a person loose if he, once and for all, declared all of his preferences to an authority and then left it to that body to execute his will, as if, so to speak, he mortified himself judicially while still being physically alive? If his declaration was sufficiently comprehensive he should apparently derive no special advantages from being judiciallly alive if he could trust the authority to be honest. Does that mean that our appreciation of freedom as a good in itself is only due to a culture.conditionned emotional attitude, being without any rational foundation?

As stated above it is possible to consider preferences as rules of choosing. A preference for coffee over tea can be given the format of a rule: *If* the agent is presented with a choice between coffee and tea *then* he will choose coffee. But even if one knows that rule, or preference, it depends on the circonstances in of the actuel situation how it shall be applied. If I am presented with a choice between coffee and tea it may be true that I normally would prefer coffee but my actual choice will depend on the context, i.e. on the situation I am in. I have no "absolute" preference for this or for that. In this sense all of my preferences, as rules of choosing, must be *defeasible*: I shall prefer coffee to tea unless there are special features of my situation I can point out as reasons for reverting my usual preference. In a state of thirst I may choose tea instead of coffee considering tea to be the better for quenching thirst; but normally the flavour of coffee is more to my liking. There may always in an actual choice be factors which the agent judge to be more important than his usual preference. So even if there exists a clear rule of preference covering "normal" cases it is an open question how and when this rule shall be applied in an actual case. As a tobacco-smoker I would love to light a cigaret some time during a long-drawn sermon in the local church. Considering the situation I shall suppress my preference.

What does it mean to have a choice to take either A or B? If A and B shall be felt as alternatives in my choosing I must be able to imagine some situations where I would choose A, other situations where I would choose B. For if I always should take A in favour of B and was utterly unable to think of cases where I should take B instead then few of us would consider A and B as alternatives in a free choice. I would feel A as something which was forced upon me and not something I had freely chosen. I might even feel exposed not to a choice but to a threat: if you do not take A you shall have B (your head chopped off). If A and B are to be conceived as true alternatives in a choice it must be felt as a problem whether to take the one or the other. I have to weigh up the different characteristics of my situation in order to find out what is best. Of course, if my preferences are concerned only with choices among bundles of *goods* it appears as an analytical truth that one will always prefer more of a good than less of it. But in ordinary language the word ”prefer” is concerned, not with goods but with actual phenomena and things. It is only when it so happens that I prefer more of a thing than less of it that I may term it a ”good”. For I do not see a thing as a good in itself. That will depend on my preferences, and on my entire situation.

In the economic theory of the consumer's choice the all-important situational characteristics are assumed to be prices and income. Other factors may be at work but they are set apart by a - not always stated- ceteris paribus clause. So in the abstract analysis only prices and income matter. We can then compute how the consumer's demands will depend on prices and income provided "all other factors" do not change. If the analysis shall pretend some realism it must be assumed that these "all other factors", i.e. the contents of the ceteris paribus clause, will refer to a normal situation for the consumer. So what the demand equations tell is how the consumer will normally behave Sometimes the very setup of the choice-problem can violate the ceteris paribus clause. If the consumer is asked to choose between living in Denmark or in France his answer may depend on whether he is in fact living here or there. "I'd rather have John's Dad than my own" says the the naive child without seeing the logical mess she may involve herself into.. So if the consumer's preferences depend crucially on whether he is in situation A or in situation B it is pointless to ask for his preferences for being in either A or B.

THE CARTESIAN AGENT

One may take these remarks as the outline of an argument against Arrow's and Debreu's idea of a "cartesian" agent who is born into the world with nothing but preferences (and reason). He prefers therefore he exists: *aliquid nihilo praefert, ergo est*. A rational cartesian agent is supposed to have preferences *ab ovo* which do not depend on the situation he is in. However, the agent's expected enjoyment may depend on the situation in which he is going to enjoy the goods he demands. But this dependency is built into the concept of a good. A good is not just a good ; it is a right to have a thing or a service delivered at a specified date, on a specified location and in a specified "state of the world". An umbrella is just an object. It may become a good by specification: as the right to have an umbrella delivered in Chicago on the first of april 2004 if it rains that day. A purchase on the spot may, as an abstraction, be conceived of as an exchange of two contracts, a promise to deliver against a promise to pay, both coming due for fulfillment within a very short time. In this way all trade which takes place in a general equillibrium model with cartesian agents will primarily be a trade in contracts, only indirectly will it be a trade in material goods and services. When trading is over each agent's dues and obligations are fixed by the contracts he has entered. If these contracts embrace all contingent (i.e.contingent on the state of the world) and dated goods, then what will be required of him and what he shall receive will be determined by these contracts for all posteriority whatever state of the world may rule. He will forever be bound like a slave to the terms of the contracts he entered in the first nanosecond of the model's existence when all equillibrium-creating trade-contracts were agreed upon by the agents. In this almost Hegelian manner freedom of contract is turned into contractual slavery in a general equillibrium model with cartesian agents. In fact, having once and for all made his decisions, about which contracts to enter, these contracts will work as an *embodiment of an all-powerful authority* that for ever in the future will decide what the agent shall receive, and what he shall do.

But even if the cartesian agent in his later life is tied, hands and feet, by the contracts he willingly entered in the initial big bang he need not feel like a slave when he, as time comes, are forced to comply with them. For he will know that compliance will give him the best possible life in relation to his preferences , abilities, and the ressources he is initially supposed to dispose of. And as the contracts are supposed to cover every conceivable contingency he cannot later argue for an annulment of his contractual obligations by a plea that things have turned out to be different from what he expected them to be when he entered the contracts. The only reason he can have for wanting to evade some of the obligations he has taken on is that his preferences in his present situation have turned out to be different from what he expected them to be when he signed. But a change of mind, and that is what this is, can never serve as a valid objection to the terms of a deal. For the whole idea of a contractual obligation is to protect parts to the contract against changes of mind that have not been foreseen by the stipulations of the contract. So if the agent shall not feel tempted to a breach of contract, or, as it may be, try the judiciary to obtain a legal evasion of his obligations, his preferences before a deal is concluded must be consistent with his preferences as they are experienced any time hereafter.

The cartesian agent makes his once-and-for-all choice according to his inborn preferences. These preferences are valid in the moment of choice; they are, however, not concerned with present enjoyments but with future contingent goods. So a cartesian agent has to know how he will pledge himself to work in the future in order to obtain an umbrella in Chicago on some day to come. He has to know how he, as a future person, would like to have that umbrella in the rainy streets of Chicago. That future person is no stranger to him: he himself will it be. But what does that mean? Noone will contend that the problem of personal identity is an easy one, certainly not in philosophy and not in practice either. I have no difficulties imagining how I shall enjoy the different goods which will go to me tomorrow, or in the next week, month or year. But when it comes to the more distant future it is more difficult for me to answer a question as to how I, as a first-person-I,would like things to be. Gradually, as the future becomes more distant I will change my first-person-I into a "third-person-I": even if I cannot tell how my preferences would be five years from now, I may be able to tell how *a person like me* might look at things. Eventually the true "myself" may disappear completely from my reasoning about what is best. For there may be situations where the only thing I can do is to try to find out "how any reasonable person would prefer things to be". Of course a cartesian agent operating in an Arrow-Debreu world of future and contingent markets may be able to pledge his entire life in the next ten years in order to ensure himself of an ample supply of Cuban cigars in any contingent future state of the world. But he cannot argue that this is what any reasonable man, even a cigar-lover, should do. For noone, not even the most reasonable person, can know if his enjoyment of a cigar will be as delectable ten years hence as it would be now.

Therefore, in order to have preferences regarding all future and contingent goods a cartesian agent must be able to imagine how he, as a future person, will enjoy the goods which will fall to him in future periods and in future states of the world. So in order to know his preferences as of now he has to know the preferences of that future person, and to know how these preferences will depend on the state of the world. Then why should his present preferences be independent of the actual state of the world when that future person and the now-person is one and the same? Perhaps the idea of a cartesian agent who is born into the world with state-independent preferences is a self-defeating conception.

PREFERENCES AND RULES

This difficulty does not appear if we consider preferences as defeasible rules of choosing. Then the problem of choice for an agent is not to find out what his "true" preferences are, but to see if there are features of his actual situation which can serve as arguments against choosing according to his normal (i.e. defeasible) preferences. If, on the other hand, preferences are seen as belonging to an inborn psycho-biological outfit it becomes difficult to understand the difference between a choice and an unconscious reaction. Did I really kick the doctor when he hit my knee with a rubber rod? There is a gulf between a willful kick and a Wassermann-reaction.

Now suppose an agent has formed his considered preferences, and that he recognizes them to be defeasible. That is he knows that in situation A he will normally, i.e. as rule, prefer to do a. However, he will not be willing to enter a contract obliging him always to do a if case A should become a reality, for there may sometimes be special features of A which will induce him to prefer not to do a but to do something else, say b. But suppose these special features can be put into the contract by explicit stipulations. Then if the agent is able to state his preferences in the format: in situation A I prefer a, in B it is b, in C c, etc, and if A,B,C,etc can be assumed to cover every thinkable eventuality the agent should have no qualms about entering a contract obliging him to do a if A ruled, b if B ruled , c if C etc. For the contract will bind him to do exactly as he would have done anyway, i.e. without a contract. So he will not find any conflicts between compliance with the contract on the one hand and his preferences on the other: it is not difficult to keep a promise to the effect that I shall do exactly as I prefer. And apparently I should not feel constrained in my liberty by the existence of an authority who just ordered me to do what I preferred to do. In fact, few people would talk of promises and orders, properly spoken, in these cases. For a promise will bind the promisor to do what he has promised even if he eventually will not like to do it. And a promise will only be felt as binding, or as a constraint on one's freedom of action, if it is possible to imagine cases where a fulfillment of the promise will not be what the promisor had preferred to do had he not been obliged by that promise. By a promise I pledge myself to do certain things as stipulated by the promise, and when the promise becomes due for fulfillment I shall have to do these things whether I like it or not. Therefore a declaration of an intent always to do as one prefers is, if not a plain truism then certainly not a promise.

However, even if I only will make a promise, or enter a contract, when the contractual stipulations, or the wording of the promise, is in accordance with my preferences, there is always a risk that the contractual specifications of my preferences turn out to be incomplete when I am required to comply. This incompleteness is due to the existence of defeasible preferences.

Why are the rules of preference defeasible? In a way defeasibility is built into the very concept of a rule. A rule can at most tell what to do in abstract cases as described by the rule,but the important thing is to find out how the rule shall be applied in an actual case. The problem for the judge is to find out to what extent the actual case resembles or disembles the abstract cases referred to in the rules of the law, and only rarely can his decision be seen as a sheer logical deduction from the rules of the law because particulars of the actual case may influence the manner in which these rules are used by the court to reach a verdict. There is no "computable" path from the rules to a decision as to how these rules shall be applied in an actual case. (If there existed such paths the entire judicial system could be replaced by a giant computer which, fed with the data of any actual case, could produce a verdict)

The logical distinction between a rule on the one hand and on the other hand the rules as to how it shall be applied in actual cases makes it impossible to conceive of "absolute" - i.e. non-defeasible -rules. For suppose it was possible to incorporate into a given rule all rules pertaining to its application in actual cases. Then it would become a problem how the ensuing super-rule, or higher order -rule, should be applied in an actual case: no rule can give a complete determination of its own validity.

THE APPLICATION OF A RULE

Now look at the rules of preference for an agent. The agent knows that if he is in state A he will prefer to do a, in state B he prefers b, in C it is c, etc. Each of these rules are defeasible. But if we can assume that the array A,B,C,etc will cover every potential state of the world then the super-rule: if the agent is in states (A,B,C,etc) he will prefer to do (a,b,c,etc.) then this rule will appear to be indefeasible, for all the judge has to do is to decide if an actual state, say X, is an A-state, or a B-state, a C,D etc.

However, to classify an observed X in categories predetermined by the rule is no easy task. Suppose for example there are four categories, A:living in a cottage,B:living in a villa, C:living in a castle, and D:living some other place and that the actual living-place is properly described by means of terms drawn from the following list:mansion, lodge, hermitage, tower, pavilion, hotel, court, manor house, hall, bungalow, log cabin, hut, hovel, shanty, shack, shed or chalet. Is a small Tudor manor with a beatiful garden in the middle of extensive wood-lands to be classified as as a cottage, a villa, a castle, or as something else? The problem for the judge is to see if an actual case, described by X, can be described by the categories assumed by the rule. An actual case can be described in many ways in ordinary language, and this is what X means: an ordinary language description. The judge's task is to translate ordinary language descriptions of the actual case into the language of the rule so that the rule can be applied to it. That presupposes the existence of rules and principles which must be abided by the conscientious judge who wants to do a proper translation. But then the problem reappears with regard to how these rules shall be applied in an actual case. So even if we assume, as we did, that the array A,B,C,etc. will cover every potential state of the world relevant for the application of a rule it can never be a matter of sheer computation to find out to which of these states an actual state, X, shall belong. The final judgment is in the nature of a decision. And so it must be for logical reasons. It can never be seen as an incontestable result of a calculation, or as an indisputable answer to a mathematical problem.

Most people will, I guess, accept this conclusion as valid in practice: in the real world court-room proceedings are not computerized, and the judge's decision is a true decision: he is the one who decides which arguments shall be considered relevant for the case and which shall not. But some may, though accepting its validity as regards practice, still feel in doubt as to its validity in principle: it must be possible to think up cases simplified to the extent that the application of a rule must appear as a result of a computation and not as a decision proper. Let us look at an example where matters are, as I see it, simplified at the utmost.

AN EXAMPLE

Suppose the states of the world can be characterized by a parameter, u, and define an A-state as a state where u is less than or equal to one, and a non-A-state as one where u is greater than one. To see if an actual state, X, is an A or a non-A we just have to measure u in X in order to see if X is characterized by an u which is greater than, less than or equal to one (think of A/nonA as sober/drunk, and of u as a parameter indicating alcohol in the blood). Call the measured value of u in X u' and its true, but unknown value u". The question is if we can make an inference from the measured value u' as to how the true value u" is situated in relation to u=1. That will depend on the accuracy of the measurement and of how close the measured u' is to u=1. The result of any actual measurement must be a rational number expressed with a finite number of decimal places. There will necessarily be uncertainty attached to the last decimal in that expression. And this uncertainty can only be removed by a more accurate measurement which can inform us on the value of the next-coming decimal. But even if it was easy, and cost-less, to increase the accuracy of the measurement to any desired degree any actual measurement would still yield a rational number with a finite number of decimals. If some purist should want a "complete" accuracy in his measurements he would have to write down decimals in one endless sequence, as if condemned to a perpetual number-writing going on from now on and not just till kingdom come but through all subsequent kingdoms as well, and, as a final damnation, he would never be able to publish the result of his "measuring". So if we accept the idea that it is possible to establish a measurement of u in X we must also accept that there is a fundamental uncertainty attached to it which will make the problem whether X is an A or a non-A mathematically undecidable if the measured u' lies sufficiently close to u=1. If u' lies in a range such that the problem is undecidable we can choose either to make a new and more accurate measurement,or we can decide that enough is enough and settle the matter by "non-rational" means, that is, by a true decision. It must be noticed that "non-rational" does not mean unreasonable. What the judge will do in such cases is to seek supplementary evidence in support of his decision. So he may have good reasons to judge X to be either an A or a non-A in spite of the fact that it cannot be determined by a measurement of u.

CONCLUSION

Now suppose an agent has a complete knowledge of his own preferences in the sense that he can make an explicit statement: in state A I prefer a, in state B it is b, in C it is c, etc, and such that the array A,B,C, etc. are mutually exclusive characterizations of every potential state of the world. What would the agent loose if he, having openly declared his preferences, left all of his decision-making to an external authority, or by entering a contract obliging him to do a in A, b in B, c in C etc.?

What he would loose would be the power to decide uncertain cases, i.e. cases where it is not a readily computable matter to decide if an actual state shall be classified as an A, a B,C etc. Then matters must be settled by decision, and if the power to decide rests with an external authority, or with a judge, there is always a risk that the judge´s ruling, although in perfect accordance with the agent's prior declaration, does not correspond to what the agent really had in mind when he sees how the judge is going to apply the rules of his declaration in actual cases.. And as this declaration is a formal(i.e. explicit) statement of the agent's own criteria for what he considers preferable, he cannot prefer to leave the interpretation of these criteria to others. In the restaurant the authority *interpretes* my preferences for different foods to the effect that I shall have fried sole to day. But the only way he can ascertain that this interpretation is correct is by asking me.

And then he ceases to be a true authority.

This can be taken as one of the elements in a proof that it is *rational for any (reasonable) agent to have a preference for personal freedom.*

This preference for personal freedom must, like preferences in general, be defeasible. What this means is that it is rational for a person to oppose any restraints on his freedom unless there are, as the case may be, a reasonable justification for constraining it. Then a preference for freedom implies, that a person will prefer to live in a system where any man-made constraint on his liberty has a justification he can understand and accept.

Literature

Immanuel Kant: *Critique of Practical Reason*, London 1996 (1787).

Robert Nozick: *The Nature of Rationality*, Princeton 1993.