

Christoph Lumer

Which Preferences Shall Be the Basis of Rational Decision?

Appeared in: Christoph Fehige; Ulla Wessels (eds.): Preferences. Berlin; New York: de Gruyter 1998. Pp. 33-56.

Abstract: Theories of rational decision normally distinguish basic and other preferences, using only the former for calculating the utility function of an agent. The idea behind the distinction is that, on the one hand, a theory of *rational* decision must allow *criticism* of at least a part of the agent's actual preferences; on the other hand, so as not to lose touch with the agent's real interests, it must rely on his *factual* preferences. Different decision theories have declared as basic various sets of preferences, thereby arriving at very different utility functions. Therefore, the question of which preferences shall be basic is of large practical importance. Nonetheless, it has rarely been discussed.

This article criticizes some standard approaches, but mainly develops criteria for the selection of basic preferences. One of the principles for the selection of basic preferences, for example, is epistemological rationalization. From these principles, then, 12 conditions of adequacy for the selection of the preferential basis are derived, e. g. taking over only intrinsic preferences, and of these not the single preferences but their underlying criteria.

Contents:

1. Introduction: Basic Preferences in the Theory of Rational Decision

1.1. Why Distinguishing Basic Preferences?

1.2. Criticism of Some Concepts of Basic Preferences in the Standard Approaches to Rational Decision Theory

1.3. The Plan of this Paper

1.4. A Terminological Remark: Subjective Desirabilities

2. The Three Principles for the Selection of Basic Preferences

2.1. Epistemological Rationalization as a Guideline for the Selection of the Basic Preferences

2.2. Adjustment to the Demands of the Whole Theory of Rational Decision

2.3. Coherence

3. Conditions of Adequacy, Group 1: Contributions to the Epistemological Rationalization

3.1. AQPBI: Only Internal Desirabilities as Preferential Basis

3.2. AQPBI2: Analytical Desirability Functions as Preferential Basis

3.3. AQPBI3: Originality of the Internal Desirability Function

3.4. AQPBI4: Wisdom of the Preferential Basis

3.5. AQPBI5: Incorporation of all Relevant Internal Desirability Functions

4. Conditions of Adequacy, Group 2: Adjustment to the Demands of the Rest of the Theory of Rational Decision

4.1. AQPBI6: Motivating Power of the Preferential Basis

4.2. AQPBI7 - AQPBI9: Quantitativity, Determination of the Desirability Function, and Splitting Condition

5. Conditions of Adequacy, Group 3: Coherence of the Preferential Basis

5.1. AQPBI0, AQPBI1: Uniqueness and Additive Coherence of the Preferential Basis

5.2. AQPBI2: Closed Circumstances

6. The Resulting Preferential Basis

1. Introduction: Basic Preferences in the Theory of Rational Decision

1.1. Why Distinguishing Basic Preferences?

Every theory of rational decision distinguishes a certain utility (or desirability) function which is basic at least relative to that theory, and an expected utility (or desirability) function which is derived from the basic utility and the subjective probability function of the agent. The basic utility function is obtained from a certain subset of the agent's preferences. This subset, which is explicitly designated by the theory, I call the "preferential basis (according to the theory)" or "basic preferences". Often the theories - given that they are searching for the basis for the expected utility - take as basic those preferences that are conceived as being independent of probabilistic considerations, e. g. preferences concerning the outcomes of an action or concerning whole worlds.¹ (But frequently, in order to obtain a quantitative weighing factor with which it would be able to measure the dependent quantity, theories take preferences which include probabilistic features as the empirical basis of the utility functions: e. g. they take lotteries or probabilistically split and combined objects (like $q \& p_1 \& p_2 \vee q \& p_1 \& \neg p_2$) as objects of the preferences forming that empirical basis.² So right from the beginning there is a certain tension, if not discrepancy, between the basic preferences and the basic utility function.)

The effect of such a division into basic preferences and other preferences is that only a *part* of the agent's preferences is adopted (without any criticism) as the basis of the utility function and thus as the basis of rational action - the other part of his preferences is neglected. The ideas behind this selection approximately are these: The theory of *rational* decision in a certain sense is a *normative* theory, suggesting which action should be chosen, so it cannot *right from the beginning* declare as rational all parts of a decision; especially it cannot declare

¹ In *Jeffrey's* theory only whole worlds (complete consistent novels) have non-probabilistic values; whereas the desirability of a proposition is a probability-weighted average of the values of the possible worlds in which it would be true (Jeffrey (1983), p. 210). In the theory of *Savage* consequences have absolute utilities, and actions get expected utilities (Savage (1972)). *Raiffa* assigns basic utilities only to the last nodes (the outcomes) of a decision tree and to special events (like paying for access to the path) on the path to these nodes; all other nodes get their utilities by an expected utility calculus over the basic utilities of the following nodes in the subtree beginning at that node and over the probabilities of the branches (Raiffa (1968), ch. 2).

² *Raiffa* takes preferences over lotteries (Raiffa (1968), ch. 4), *Jeffrey* objects of the form $q \& p_1 \& p_2 \vee q \& p_1 \& \neg p_2 \vee q \& \neg p_1 \& p_2$, with q as the measured object and p_i as events like obtaining heads as a result of tossing a coin (Jeffrey (1983), pp. 134-136), as objects of the basic preferences.

per definitionem as rational the decision in a narrower sense, i. e. the final preference for one of the alternative actions which *actually* leads to the realization of this action. There must be the possibility of *criticism*; so the theory cannot accept all of the agents thoughts as infallible. On the other hand, the theory of rational decision has a certain scope: It must be in the best interest of the agents to apply this theory because its application guarantees the maximal realisation of their desires and interests. Thus the theory of rational decision cannot be independent of the agents' *factual* desires and preferences, it must adopt at least some of them as the uncriticised basis of its constructions, and it must *maximize* the realisation of *these*. Another point is that, if the theory were not to take into consideration these factual desires, no one would apply it and act on its recommendations, so the theory would be practically useless. Therefore, something like a differentiation into basic and other preferences, and the adoption exactly of the former of these, is necessary for any theory of rational decision. And the basic preferences should be more natural or have, at least, less epistemic ingredients or elements, such as probabilistic considerations, so that the choice of them rather than others as basic, and not to be criticized, would be justified.

1.2. Criticism of Some Concepts of Basic Preferences in the Standard Approaches to Rational Decision Theory

In the various approaches to the theory of rational decision very different sets of preferences are taken as basic (see footnote 1). But these differences also lead to different (expected) utility functions, and to different preference orders. This results from the fact that, normally, the agents' preferences are incoherent (this was one motive for introducing a theory of rational decision); so if one begins with one subset of preferences in order to determine the underlying utility function, and then coherently calculates what the rest of the preference order must be, one arrives at rational preferences which differ from those of the agent; and starting with another subset of preferences one arrives at still another preference order. Therefore also in the very practical issue of which action should be chosen rationally, the question of which subset of preferences is taken as the basic one, is extremely important. However, in the relevant literature the problem of how to confine the basic preferences is hardly discussed, and frequently treated as a merely technical question of which preferences would be the best basis for measuring utilities. Or the question, within certain limits, remains open to the pleasure of the user.³ The resulting determinations of the basic preferences are quite

³ E. g. the very instructive survey reader *Gärdenfors and Sahlin* (1988) does not contain any contribution on this subject. *Jeffrey* sets aside the difference between desirability and estimated desirability for normal objects of a valuation, and considers all desirabilities as estimated ones; the argument for this view is that any object *a* of valuation can be split up into the parts *a&b* and *a&-b* from which the estimated desirability of *a* can be calculated by means of the probability-weighted average (Jeffrey (1983), p. 87). Only if the objects are whole worlds is further splitting no longer possible, so that one is entitled to speak of "desirabilities" (Jeffrey (1983), p. 210). *Raiffa* leaves open what are the finest ramifications of his trees; but in his examples the ends of the ramifications are mainly amounts of money and lotteries over them (Raiffa (1968), ch. 4).

inadequate.

The following are some of the main problems concerning the basic preferences in standard approaches to rational decision theory:

- a. If the distinction between basic and other preferences is discounted completely, or at least in the practice of decision, and if it is left open to the user's pleasure from which factual preferences he starts his rational deliberation, then this, according to the above explained argument of incoherence, will result in an *arbitrariness* of the rational actions themselves. If the user does not want mere coherence, but *rationality* of his decisions, he himself must look for suitable restrictions, without the help of any theory; i. e. decision theory as theory of *rational* decision then has not accomplished its task.
- b. If in the theory there is a clear-cut difference between basic utilities and expected utilities, then frequently the preferences used for *measuring* these utilities are not preferences over the objects for which *basic* utilities are defined.⁴ Hence *the theory does not provide any means for measuring basic utilities*; thus, also, the expected utilities dependent upon them are not determinable.
- c. The preferences taken as basic are *far from being natural* or without epistemic elements; on the contrary, they are quite sophisticated, resting on probabilistic 'calculations' and on far-reaching empirical knowledge, and summing up long chains of desirability values. (What would you prefer: a 0.02 lottery for the atomic warfare (i. e. a lottery, where with a chance of 0.02 the result will be the warfare, and with a chance of 0.98 that, afterwards, nothing will happen (cf. Raiffa (1968), ch. 4)) or a forfeit of \$5,000,000? Before answering such questions seriously, we think carefully about the consequences of our choice, and the various probabilities.) Why should such preferences be excluded from rational criticism, while preferences among actions are not? If those preferences, as I have argued, are far from being really basic, decision theory right from the beginning could take preferences over *actions* as "basic" - and thereby make itself superfluous.
- d. Because the "basic" (according to the theory) preferences already rely on complex calculations, it is not too surprising that they are *all* quite *incoherent*. The problem with incoherent preferences is that the theorist does not now know which *part* of them he should take as the basis of the utility function. One answer to the problem is to introduce a nonlinear preference and utility theory (cf. e. g. Fishburn (1988)). But is this the right path? Is it not, in effect, an attempt to stick to a basis which is already a product of extensive reflection and therefore full of miscalculations, as well as an attempt to declare all this to be unimpeachable?
- e. There is a long theoretical tradition regarding the *criticism of desires*, and implicitly also of the preferences which rest on these desires: External criticism measures the desires e.

⁴ This is the case e. g. in the theory of Jeffrey: The objects of the basic preferences are propositions of the form $q \& p_1 \& p_2 \vee q \& p_1 \& \neg p_2 \vee q \& \neg p_1 \& p_2$; basic, non-probabilistic desirabilities are assigned only to complete worlds (Jeffrey (1983), pp. 134-136; 210).

- g. by moral standards, standards of perfection or standards of a certain ideology. For an internal criticism there are several approaches: Fulfilling the desire may not serve the underlying purposes and therefore not be in the real interest of the agent; fulfilling the desire may not lead to satisfaction; the desire or preference may not motivate the subject to the corresponding action; and so on (cf.: Kusser (1989)). Such criticisms entail independent parts of standards of rationality, and these requirements of rationality are completely disregarded when utility is based on the standard conceptions of basic preferences.
- f. Taking factual (i. e. credited, or conscious, or uttered) preferences directly as the basis of utility functions has the consequence that these functions will be *extremely incomplete*. What is the utility of an object p , over which the subject has never formed a (real) preference? Thus p may be part of an alternative never executed, but also a state of affairs which we would estimate, when being asked if we do, but which we never did rank in any way. One could define the utility of p by means of *conditionals* (the utility of p is determined by the preference over p the subject would state if asked to compare p with q); but then we require a strong theory upon the development of preferences, a theory which perhaps would do away with relying on single preferences. Another problem with hypothetical questions for, and utterances of, preferences is that the question and the utterance themselves may influence the object of evaluation ("During the concert we were disturbed every five minutes by this annoying utility researcher ...").
- Perhaps the most important criticisms are (c) and (e). These criticisms imply that the preferences chosen as basic by the standard theories of rational decision are neither basic (epistemically or from the perspective of the subject) nor rational (in the sense of withstanding criticism). So the theory of rational decision urgently needs an enlargement with the aim of determining the adequate basic preferences.⁵

1.3. The Plan of this Paper

The aim of this paper is to present an approach for this part of the theory of rational decision. In this approach, the *intrinsic desirability* will be the basic desirability, with which at a later stage will be defined other concepts of desirability. This approach can be called the "*enlightened motivational theory of the intrinsic good*". Here I will expound the three *principal ideas* of this approach for singling out basic preferences (2.) and then develop *conditions of adequacy* for the preferential basis of the intrinsic desirability; these conditions of adequacy are subdivided into three groups, according to the underlying principle on which they are based (3.-5.). The development of the definition of 'intrinsic desirability' itself is rather extensive and technical and is therefore here omitted. The final section gives some hints what empirically might remain, after applying the conditions of adequacy, as the preferential basis of rational decision (6.).

⁵ I also think that many of the difficulties which have arisen in the attempts to develop rational or game theoretical ethics stem from here: The preferences taken as basis for these attempts are quite irrational.

1.4. A Terminological Remark: Subjective Desirabilities

Before beginning with the systematic part of this paper I must make some remarks regarding terminology. The original idea behind the standard method of defining basic utilities by referring to *preferences* is to have a clear-cut *behavioral* and therefore "empirical" basis: Certain alternatives are offered to the subject, and he or she answers with a choice, ideally with an action. But this behavioral idea is today mere fiction, because the things to be chosen according to the questionnaire of the utility researcher are only hypothetical actions ("What would you prefer, a 0.25 lottery over \$ 10,000 or ...?") or are not actions at all ("What would you prefer, i. getting a good job and heads in the first and second toss or getting a good job and heads in the first and tails in the second toss or ii. ...?"), and the answer is only a verbal one: "I would prefer p to q ." And this now does not mean more than: "I like/estimate p more than q / value p higher than q ." This is the expression of a comparative *proposition on subjective valuations*. If the statement is honest and truthful there must be a corresponding *belief* of the subject about the value or desirability of p and q .⁶ This belief is comparative too, but it may be founded in vague quantitative ideas of the desirability of p and the desirability of q . Of course, these propositions believed in are not really quantitative; but they are more than comparative in that they e. g. range p and q in vaguely envisaged intervals of desirabilities. Therefore there is no harm done if, instead of ' s prefers p to q ', one uses the concept of *subjective desirabilities*: ' s believes that the desirability of p is quite high'. Because this is a very much richer concept I will often use it instead of speaking of preferences.

2. The Three Principles for the Selection of Basic Preferences

There are three principles from which the conditions of adequacy for the preferential basis of the intrinsic desirability are developed. These principles are expounded in the following three subsections.

2.1. Epistemological Rationalization as a Guideline for the Selection of the Basic Preferences

What does it mean to decide rationally? The most simple answer to this question is: To decide rationally means to choose the best action. But this answer is empty if one is at the beginning of a theory of rational decision, and still has to define intrinsic, total and expected desirabilities. So we have to look for answers which are more informative in the sense that they expose the ideas behind our definitions of rational decision, the ideas of practical rationality from which are developed the rules of a well-defined rational decision.

⁶ Here I can leave undecided whether or not this belief itself is already a desire, or the core of a desire, or if it is only a cognitive representation of a desire. For the theory exposed here the only important thing is that such beliefs are connected with a corresponding "motivating force", and it is unimportant by which intermediate processes this connection is established.

Every decision is a result of the interplay of motivational forces and beliefs; and the motivational forces too can be analyzed into components of pure propensity and components of belief. The principal idea of every theory of rational action or rational decision is to introduce *epistemological* rationality into the believing components of decision. Elements of *epistemological rationality* e. g. are: not simply to believe, but to believe on epistemologically well founded grounds (a justified belief is probably nearer the truth); to abandon a belief as soon as another, incompatible with it, becomes more reasonable; to extend one's knowledge. The introduction of epistemological rationality into the believing components of decision thus could take the following form: expanding the relevant information on which the decision is based ("relevant" here means that such information would alter the decision); or excluding false beliefs or background convictions on which the decision could be based. Such ways of introducing epistemological rationality I term "epistemological rationalization of decision".

The reasons behind the approach of envisaging *practical* rationality as *epistemological* rationalization of the decisions are these: Among the two components of decision, beliefs and pure propensities, we can cognitively influence only the beliefs, and about beliefs we know rather clearly from epistemology what it means that they are "rational"; so the epistemological rationalization of decision is the *only* way to practical rationality.

The principal idea of introducing epistemological rationality into the believing components of decision is not new ⁷ and - at least in a vague form - common to all theories of practical rationality and of rational decision. The special approach presented here will consist only in radicalizing this idea, and exploiting hitherto unexhausted possibilities of epistemological rationalization.

How can the idea of epistemological rationalization be applied to the selection of the preferential basis of intrinsic desirability? There are two directions for such an application. 1. One aim of the differentiation into basic and other preferences is to neutralize cognitive errors in the utility functions. That is, some preferences may be cognitively *derived* (in a logical sense) from other preferences or underlying criteria, and some empirical information. During this deduction errors may occur. The only subjective preferences, desirabilities or criteria which are definitely free from such cognitive errors are the real *fundamental* ones, those which are not derived from underlying criteria or subjective desirabilities. So if the preferences taken as basis for the utility functions are really fundamental, they cannot contain cognitive errors. The *utility theory* then tells us by means of its definitions how to derive correctly the desirabilities and non-basic preferences from those basic preferences etc. Therefore, the basis of the desirabilities should be *really fundamental* in the sense of not being derived from underlying criteria, preferences or subjective desirabilities. 2. Preferences, subjective desirabilities or criteria, though not being *derived* from other preferences etc. and empirical information, may still be *stimulated* by some new knowledge. I. e. you may have a

⁷ This idea of epistemological rationalization is already the main idea in Plato's (Charmides 166) and Aristotle's (Eth. Nic. 1140a-b) concept of prudence (phronesis).

really fundamental subjective desirability; then you experience something new or obtain fresh knowledge; and this causes you to change your subjective desirability because you think the new desirability is more appropriate or adequate, or something similar, but you cannot say why; there is no real epistemological reason for changing your subjective desirability. The new subjective desirability is also fundamental, because it is not derived from the old one and the new knowledge, but is *based* on (but not epistemologically *justified* by) a wider store of knowledge; and in this regard it is *wiser* (resulting from greater wisdom) than the old subjective desirability. In such cases the idea of epistemological rationalization would require to take as the basis the wisest of the really fundamental subjective desirabilities, or criteria for them.

So the general line for application of the idea of epistemological rationalization to the differentiation of basic and other preferences is this: 1. All the steps or elements of a decision that can be reconstructed purely epistemologically (like justifications, or derivations of new knowledge or of new criteria from older sources) shall be reconstructed epistemologically *in the theory* of rational decision and not adopted from the subject as given. For such steps the *theory* defines exact rules. The aim of this procedure is to rule out possible errors of the subject, and also to show how such errors can be avoided. The theory will only adopt from the subject elements and steps which cannot be reconstructed purely epistemologically. 2. Of such elements the theory adopts the wisest (in the special sense explained above).

2.2. Adjustment to the Demands of the Whole Theory of Rational Decision

The selection of the preferential basis is only *one part* of the theory of rational decision. This preferential basis has to fulfill a certain function in the complete theory: Relying on this basis we must define the intrinsic and the total desirability of states of affairs and of the expected desirability at least of an action. These definitions must be such that subjects under certain circumstances are willing to decide between alternative actions according to the criterion of maximal expected utility, and so on (the relevant list of requirements will be presented below). Of course the preferential basis of rational decision must be chosen in such a manner that it can fulfill these requirements of the whole strategy of the theory of rational decision.

2.3. Coherence

"Coherence" generally means 1. that propositions or preferences are connected to such a degree that some of them can be derived from others, and 2. that their union is consistent, especially that the result of the derivations is not contradictory to an original preference or proposition. The adjustment to the requirements of the whole theory of rational decision will afterwards make it necessary for the preferential basis of rational decision to be interconnected in a certain way (axiological splitting of overlapping objects, see below AQP9); so that here we have only to regard the second condition of coherence: the consistency.

The preferential basis of rational decision must be consistent for the following reasons: If inconsistent, the expected desirability of an action will be different, depending on which part of the inconsistent basis is used to calculate the expected desirability. The consequence is that no further definite recommendations on which action is to be taken can be given on the grounds of the theory of rational decision. The theory then recommends *several* mutually exclusive alternatives and - even worse - at the same time asserts for each one of them that it is sub-optimal and therefore not worth choosing.

In the following sections conditions of adequacy for the preferential basis of intrinsic desirability will be presented. These conditions state *which* preferential basis shall be used for the determination of the intrinsic desirability, and in *what manner*. In some cases, conditions of adequacy for the definition of the 'intrinsic desirability' itself will be introduced. These conditions, then, represent the demands with which the preferential basis must comply (see above, sect. 2.2.).

3. Conditions of Adequacy, Group 1: Contributions to the Epistemological Rationalization

3.1. AQPBI: Only Internal Desirabilities as Preferential Basis

Here the expressions "*intrinsic desirability*", "*extrinsic desirability*", "*total desirability*" and "*expected desirability*" are used as *technical* terms, which have to be defined with complete technical dexterity. For denoting the appertaining concepts of our everyday thinking I will use expressions more frequently occurring in ordinary language: "*internal desirability*" for denoting the underlying idea of intrinsic desirability, "*external desirability*" for denoting the idea of the extrinsic desirability, "*plain desirability*" for denoting the concept which is part of the agent's decisive judgements on alternative actions, and which especially can - but must not - be a simplified equivalent to the concept of 'expected desirability'. Internal, external and plain desirabilities are not really quantitative, but are preliminary forms of quantities (in the sense explained above), whereas intrinsic, extrinsic, total and expected desirabilities are really quantitative. *All* these concepts here are to be considered as *relative to a subject* (the subject of value). The *objects* of desirabilities are always *states of affairs*, especially events and states.

Internal desirabilities are ascribed to states of affairs *per se*, and not because they have certain consequences. External desirabilities, on the other hand, are ascribed to a state of affairs because it has consequences which, in turn, have an internal desirability. So the objects with internal desirabilities are the ones in which we are essentially interested. The notion of 'internal desirability' is cognate with that of an 'ultimate end': In both cases there is no further end (or negative state) to which this end is only a means (or the cause). The condition for internal desirabilities is meant here in a rather strict and narrow sense, so that often very careful reflection will be necessary to discover if a subjective desirability is really ascribed to

an object for its own sake. One way to do this is to alter hypothetically the normal consequences of the object by introducing extraordinary circumstances and seeing if the subjective desirability remains unaffected by this. E. g. according to my subjective desirability functions, eating a good meal has no internal desirability at all, only the actual tasting of the food, the feeling of pleasure experienced while eating such a good meal and enjoying the nice surroundings etc. have (positive) internal desirabilities. I infer that eating itself has for me no internal desirability because, for example, I will not be able to appreciate my favorite dish when feeling ill, as the illness together with the actual eating would result only in a stale taste and may also cause me to feel nauseous.

In this example the valuation of the object "eating" is shown to be completely *dependent on the consequences* by hypothetically introducing extraordinary circumstances, but this does not mean that internal desirabilities must be independent of the *circumstances* - so long as this dependency is not implied by the fact that the desirability of the object is dependent on its *consequences*. In fact, internal desirabilities *can* be dependent on the circumstances, and sometimes are: E. g. for me personally, my being in a good mood has a positive internal desirability; but this desirability differs according to whether or not this good mood has occurred naturally, or if it has occurred as a result of taking drugs; in the latter case the desirability is lower. But in this example it is not the fact that the circumstances together with the (same) object of valuation (i. e. good mood) may cause different *consequences* which leads me to estimate the object in a different way. Perhaps the altered circumstances lead to different consequences, e. g. to a hangover, but perhaps they do not; the devaluation is quite independent of this: I have experienced the good mood in any case, and this internally is more or less good, depending on whether or not induced by drugs but not depending on the effects; what will be altered by the hangover effect is the *external or plain* desirability of *my taking drugs*. - If the internal desirability is dependent on the circumstances it is a 3-adic function: 'the desirability of the state of affairs p under the circumstance c for the subject s ', otherwise a binary one: 'the desirability of p for s '.

(Against the differentiation of the object and the circumstances of the internal desirability one may raise the objection that this is a superfluous complication which can be avoided by taking the circumstances as part of the object. Then, being in a good mood and having taken drugs is another object than being in a good mood and not having taken drugs; and it stands to reason that different objects may be valued differently. But this suggestion does not help here because the object and the circumstances in our calculations really are treated differently. Especially, the internal desirabilities of different *objects* p and q may simply be *added* to the desirability of $p&q$ only if p and q do not overlap; but their *circumstances* c and d may overlap. E. g. let p be my being in a good mood at a given time t , q my having an intensive and pleasant taste of caviar at t , and c and d both be my being under the influence of LSD at t ; then p and q do not overlap, but c and d do; and the internal desirability of $p&q$ under the circumstance $c&d$ equals the sum of the internal desirabilities of

p under the circumstance c , and of q under the circumstance d .)

External desirabilities clearly are *derived* from 1. an empirical proposition about the relation between the object p and a consequence q , 2. the internal desirability of q and 3. the definition of external desirabilities. *Plain desirabilities* are derived in an even more complicated way. Thus external and plain desirabilities are derived from other desirabilities, and, depending especially on the empirical premise (1.) of their derivation, they may be false. Only the internal desirabilities are perhaps not derived in this way. Therefore, according to the idea of excluding possibilities of error (see above, sect. 2.1.), the first condition of adequacy for the preferential basis must be:

AQPB1: Subjective internal desirabilities as preferential basis: *Only subjective internal desirabilities (in the strict sense expounded above) may be the preferential basis for the intrinsic desirability.*

3.2. AQPB2: Analytical Desirability Functions as Preferential Basis

Internal valuations are not performed arbitrarily, but according to (implicit) *criteria*. Such criteria are normally not conscious, and they can be read only in the person's inward valuing "behavior": She or he detects a certain *property* of the object of valuation and *therefore* ascribes a certain internal desirability to it, and the relation of substantiation indicated by the expression "therefore" is *not* one of an *external* desirability; i. e. detecting the relevant property is the *only* reason for ascribing the internal desirability. The fact that the agent makes evaluations according to (implicit) criteria implies that there is a *general* rule or *analytical desirability function* Φ to which he tries to orientate his valuations. So the inward valuing "behavior" consists in events of the following type: The subject s detects that an object p has a certain Φ -value of u_p , then s ascribes to p the internal desirability of u_p ; the detection of the Φ -value is the reason for ascribing the desirability, and there is no other reason for this ascription (except perhaps the belief that if an object has the Φ -value x it must have the internal desirability of x).

If there are such criteria for internal valuations (and I have just said that there are) then the single internal valuations in any case are *derived* from the criterion and the proposition that the object does fulfill it. Therefore the single valuation can be *false*, e. g. the person may be wrong in thinking that p has the Φ -value u_p . And according to the idea of fundamentality (see above, sect. 2.1.) *single* subjective internal valuations as such should not be the preferential basis for the intrinsic desirability. However, primary to the single valuation, and perhaps not derived from other preferential items, is the *criterion* for, or *general analytical function* of, the subjective internal desirability. Therefore, only *general analytical functions* of the subjective internal desirability may be the preferential basis of the intrinsic desirability. A positive by-product of this approach is that the holes in the intrinsic desirability function stemming from the single preference approach (cf. above, sect. 1.2., critique f.) are now filled: If we know the general (analytical) function of the internal desirability, the internal

desirability of each of the objects internally relevant to the subject is defined, even for those the subject has never thought of.

Some of the general analytical desirability functions are applied constantly during longer periods; but perhaps after a certain time or because of certain experiences they are abandoned, and replaced by others. There may even be general analytical desirability functions the agent *would* adopt if he had certain knowledge which he actually has not and perhaps never will have, e. g. because his intellectual development will be interrupted by an early death. Such mere *hypothetical* general desirability functions too shall be elements of the set from which the preferential basis is chosen. The reason for this widening of the relevant set of desirability functions is this: According to the idea of epistemological rationalization (see above, sect. 2.1.) the epistemological apparatus afterwards will be used *destructively* to rule out certain desirability functions as not being rational, or something like that. However widening the set of relevant desirability functions in the described manner means introducing the *constructive* power of epistemological rationalization in the strongest form: A certain and even hypothetical knowledge may induce new desirability functions. Moreover, the epistemological critique may be so sharp that it rules out *all* the *factual* desirability functions of the agent, so that without the discussed widening there would not remain *any* preferential basis for the intrinsic desirability at all. Therefore the power of additional knowledge is a necessary counteracting force to the epistemological critique.

So the second condition of adequacy is:

AQPB2: Analytical internal desirability functions as preferential basis: *The intrinsic desirability for a subject s must be based directly on factual or hypothetic knowledge-induced analytical internal desirability functions of s in such a manner that it is covariant with them, having the same (preferential) ordering as exists between the sums of internal desirabilities; i. e. the intrinsic desirability is a positive-linear transformation of the quantitatively interpreted analytical internal desirability function or of the sum of several of such functions. A factual desirability function of s is one s applies for some time in his life; a hypothetic knowledge-induced desirability function of s is one s would adopt if he had a certain knowledge.*

3.3. AQPB3: Originality of the Internal Desirability Function

With an (analytical) internal desirability function and suitable empirical assumptions *external* or *secondary internal* desirability functions can be established; e. g. the (internally irrelevant) objects of the type F can always cause objects of the type G , which have an internal desirability of u_g ; therefore all the F have an *external* desirability of u_g ; or if an object is F (which is internally irrelevant) it is also G ; and being G is internally desirable to the degree u_g ; therefore all the F have a *secondary internal* desirability of u_g . Such external or secondary internal desirability functions help to save certain steps of the deliberation: If the agent detects that an object is F he must not consider that the F are G ..., but immediately can infer that it is (externally or internally) desirable to the degree u_g . But such external or secondary internal

desirability functions can *make themselves independent*: The agent uses them mechanically without keeping their origin in mind; after a while he has forgotten or even repressed that they are only external or secondary internal desirability functions, and he takes them as original internal desirability functions.

An external or secondary internal desirability function becoming independent is innocuous if the originally assumed empirical relation between the F and the G really exists; but this may not be the case. Typical examples for such secondary internal desirability functions relying on false beliefs are: 'If mummy says x is good for me, then x is internally good for me.'; 'God's will is internally good for everyone.' Sometimes it will be rather difficult and perhaps require psychoanalytical means to find out if a, now internal, desirability function perhaps is not, originally, an external or secondary internal one.

External or secondary internal desirability functions having become independent are surely not fundamental (in the sense explained above, see sect. 2.1.). Therefore, according to the strategy of an epistemological rationalization only originally internal desirability functions can be the preferential basis of the intrinsic desirability.

AQPB3: Originality of the internal desirability function: *Only originally internal desirability functions may be the basis of the intrinsic desirability.*

Against these first three conditions of adequacy it might be objected that subjective internal desirabilities change too often; therefore, they cannot constitute a stable and really fundamental basis of rational decision. Surely, if one asks a person "what is internally good / bad for you?" one might get very different answers, depending on the time of asking. Also, the conscious ends of action may change quickly. But these things are not the preferential basis intended by my theory: According to the criteria presented so far, not the *conscious ends* of action are relevant, but the last underlying practical reasons the agent could present after intensive reflection: Why did you want this? Because of p . And why did you want p ? And so on. Then it is not the *single evaluation* that is relevant, but the underlying implicit criterion. And of these it is not those that the agent actually uses as internal criteria that are relevant, but the (perhaps ascertainable only by the means of depth psychology) original ones that are relevant. These three conditions already are rather narrow. Due to the absence of appropriate scientific research (cf. e. g. the survey given by Heckhausen (1989), pp. 455-466) we have no statistical data which could sustain my view; but I would estimate that only very few, if any, people change their original criteria for internal desirability more than ten times in the course of a lifetime. And the following conditions of adequacy (above all AQPB4, AQPB6 and AQPB12) will reduce this set of maximally ten criteria once again to one or two criteria. But these are already *empirical* speculations, which shall be continued in section 6.

3.4. AQPB4: Wisdom of the Preferential Basis

If one develops an external or a secondary internal desirability function there is initially no intention to change the basic internal desirability function. But a real change of it is *intended*

in another case: As already mentioned, the agent is normally not conscious of the criterion for the internal desirability, he only detects certain properties of the object and then values the object to a certain degree. According to this type of materialization of the criteria for the internal desirability, it is possible that the agent one day detects *new* properties of the object, which spontaneously seem to him to be relevant, even though so far disregarded by him, and which motivate him to alter his mode of valuation: He values the object to a different degree than he would have done without his discovery, i. e. according to a *new analytical internal desirability function*; and this is only the beginning of a general change: Henceforth when internally valuing, at least in relevant cases, he watches out for that property, and when he detects it he uses the new mode of valuation.

An *example* of such a development would be this: The agent hitherto has used a simple hedonistic desirability function; but then he thinks of the possibility of getting happy by perfect drugs: By them he will feel really happy and there will not be any hangover afterwards, both of which, according to the hedonistic desirability function, he should appreciate very much - but he does not. Perhaps he feels a sort of aversion to this type of manipulation, or something like that. In any case, although not valuing such experiences negatively or neutrally, he *depreciates* them compared with the internal desirability according to the simple hedonistic function.

In such cases the new internal desirability function has a predecessor and it is *stimulated* by some new knowledge, but the new function is not *derived* from the old one and that knowledge. The new function is not implied by the old; it is, rather, a real *further development* of the old one. And because this development is inspired by greater knowledge, the new desirability function in a certain sense is *wiser* than the old. If there is a development of internal desirability functions to greater wisdom, then the idea of epistemological rationalization (see above, sect. 2.1.) requires us to take as preferential basis the wisest of them.

AQPB4: Wisdom of the preferential basis: *If there is a chain of internal desirability functions used by the agent one after the other, so that the successors inspired by new knowledge emerge from their predecessors, but (according to the agent's intention) are not derived from them, then only the ultimate, i. e. the wisest, link of this chain may be the preferential basis for the intrinsic desirability.*

Against this demand for wisdom one might object that by it the internal desirability of *suspense* would be altered in an inadequate way: Receiving information on e. g. the exact content of a lottery ticket before buying it, the possible buyer will change his behaviour: He will buy the ticket if it is a winning ticket, not buy it if it is a blank, and in neither case will he feel any suspense. This is all true, but it is no objection to the demand for wisdom. What is changed in such cases by the additional information are 1. the *plain* evaluation of the act of buying the lottery ticket and 2. the psychic state of tension, but not the *internal evaluation of the suspense*.

3.5. AQP5: Incorporation of all Relevant Internal Desirability Functions

I think that actually there will remain only *one* subjective desirability function fulfilling all the other conditions of adequacy for being the preferential basis for the intrinsic desirability, namely a corrected hedonistic one. But, at least theoretically, *several* may remain. Their having passed all the conditions means that there are no objections to any of them, and that any of them is suitable for forming the preferential basis of the intrinsic desirability. Moreover (according to AQP4) they must be independent of each other and all have their own right. In such a case *all* the relevant desirability functions should be incorporated in the intrinsic desirability.

AQP5: Incorporation of all relevant internal desirability functions: *If there are several relevant desirability functions fulfilling all the other conditions of adequacy for being the preferential basis for the intrinsic desirability, then these functions have to be added, and the summational function is the preferential basis.*

4. Conditions of Adequacy, Group 2: Adjustment to the Demands of the Rest of the Theory of Rational Decision

Singling out the preferential basis of the intrinsic desirability is only a small part of the theory of rational decision; and that singling out must fit to the whole strategy of the theory and especially to the desired properties of the intrinsic desirability. Here is not the place to expound my strategy for a theory of rational decision, so the exposition of the two requirements for the definition of the 'intrinsic desirability' which are relevant in our context must do.

1. The concept of intrinsic desirability is mainly used to define the 'total ...' and the 'expected desirability'. Both are quantitative concepts. The concept 'total desirability of p for s ' serves to measure the amount of intrinsic desirability (for s) of all the objects 'accompanying' in a certain way p ; the exact definition of 'accompany' here can remain open. The amount of intrinsic desirability of all the objects is a sum of intrinsic desirabilities of a set of objects with the following features: the set must, in a certain way, include *all* the objects having an intrinsic desirability and accompanying p ; but these objects in pairs may *not overlap* like the objects $p&q$ and $q&r$ would overlap; otherwise the overlapping parts would be counted twice. - The *expected desirability* is just something like the *expected value* of the total desirability.

2. The concepts of total and expected desirability must be the key concepts of a *practical* deliberation. I. e. if the subject s is convinced that an alternative a_1 among all the actual (able to be performed in the next moment) alternative actions taken into consideration has the highest total or expected desirability, then s is (*directly*) *motivated* to perform a_1 : If he really is able to perform a_1 , has no strong unconscious desires ... he will really do a_1 .

4.1. AQP6: Motivating Power of the Preferential Basis

If the beliefs about the internal desirability are to be used to establish beliefs on the total or expected desirability, and certain of the latter beliefs must be directly motivating (in the sense just explained) then the relevant internal desirability functions must be *indirectly motivating*: The agent must use judgements according to this desirability function to know directly motivating judgements on the highest plain utility of an actual alternative; this knowing must be direct too: the internal desirability must be taken directly as part of the plain utility.

It is *not* a matter of course that an internal desirability function is indirectly motivating: There may be internal desirability functions which are not even part of the action system, namely desirability functions according to which those value judgements that engender our affects are formed.⁸

AQP6: Motivating power of the preferential basis: *The desirability functions taken as preferential basis for the intrinsic desirability must be indirectly motivating: If the agent had a certain knowledge, he would use judgements formed according to this desirability function to know components of directly motivating judgements on the highest plain desirability of an actual alternative.*

It has been argued against my theory that internal desirabilities were socially induced into the subjects by education, and that behind the social values might have stood consequentialistic criteria, so that the "internal" evaluation would have been based upon external evaluation. Therefore, not only the difference between internal and external evaluations would have vanished, but also epistemic errors could have slipped into the internal evaluations - contrary to the idea of fundamentality. This objection disregards the *practical* intentions of the concept of internal desirability in a rational decision theory and the fact that, due to these intentions, AQP6 demands that only (indirectly) *motivating* internal desirability functions may be the basis of the intrinsic desirability. These functions do not constitute *any* opinions, which one would simply adopt from other people; instead, they are parts of our motivational systems and represent the core of our individual *will*. Everyone who has to do with children knows that already small children have their own and rather strong wills. In order to get them to *do* something *we* want we must rely on something they internally desire or refuse. (This is the role of a sophisticated system of rewards and punishments and of explanations which make the child believe in the relation between certain acts and events and his or her internal desirabilities.) Therefore, education *can* induce motivating *secondary* internal desirability functions (or perhaps originally internal "desirability" functions in another than the practical sense of "desirability") to children; or it can stimulate the transition to a wiser desirability function. But education cannot freely induce motivating originally

⁸ The relation between affects and actions is rather complicated. I think both are caused by value judgements, but by value judgements of a different kind. At least these value judgements *must* not be of the same type.

internal desirability functions to them which are not already (at least rudimentally) parts of their motivational system.⁹

4.2. AQP7 - AQP9: Quantitativity, Determination of the Desirability Function, and Splitting Condition

To be suitable for defining the total and the expected desirability in the above mentioned way 'intrinsic desirability', firstly, must be a *quantitative* notion. Secondly, the objects to which an intrinsic desirability is related must be *determined* in the way that it is defined whether or not two of them overlap and in what the overlapping or not overlapping parts consist. Thirdly, if the intrinsic desirability is defined for two overlapping objects then it must also be defined for the overlapping part and the one or two not overlapping parts (= *splitting condition*). The third condition being fulfilled, it is rather easy to solve the problem of overlapping mentioned above: If the intrinsic desirability is defined for $p&q$ and $q&r$ then according to the splitting condition it is defined also for p , q and r , so that the total amount of intrinsic desirability of the objects $p&q$ and $q&r$ together may be equated with the sum e. g. of the intrinsic desirabilities of p and of $q&r$; since the latter two objects contain all intrinsically relevant parts of $p&q$ and $q&r$ and do not overlap.

These are conditions for the *intrinsic* desirability. The *internal* desirability functions taken as the preferential basis for the intrinsic desirability must, rather, be such that (at least with suitable technical help) they can be interpreted as being quantitative, determined and fulfilling the splitting condition. I. e. if a desirability function does not *immediately* fulfill these conditions there must be an *interpretation* touching it up to the required degree but not garbling the underlying intention of its use.

Especially the quantitativity condition for the internal desirability functions, even if weakened in the sense just explained, may seem to be too strong so that no really adopted function would fulfill it: In our natural deliberations we do not use numerals as arguments of

⁹ Not to pay attention to the motivational side of practical desirabilities is also the main fault of value objectivism, which holds that value judgements are based on value-intuitions about value-properties. The first objection against this theory will always be an *epistemological* one: We do not have (something like) a sense for detecting value-properties. But the main objection against this theory is a *practical* one: Even if we had such a sense for values and if we did ascribe "value-properties" to objects according to our value-feelings, all this could not *constitute* the (motivational) desirability of an object. This is because the desirability meant in the theory of rational decision is a *motivational* or *practical* relation between an intended object of valuation and the *actions* of the subject: The subject may ascribe to the object value points or other quantities, but however objectively this is done, the result cannot alone constitute the desirability. What transforms the judgement into the relevant subjective *desirability* is the fact that the assumed quantities actually play a certain role in the practical deliberation or are destined to play such a role; namely, in a decision they must be *weights* for those actions which are accompanied by the appertaining objects. Such transformation of assumed quantities into practical weights objectively is a mere empirical fact. From our *internal* perspective it is a rationally unjustifiable intuition.

the internal desirability. But this is not actually required by that condition: If an internal desirability judgement logically has the form 'The internal desirability of p for s under the circumstance c is u_p !', firstly, ' u_p ' must not be a *numeral*, but can be a definite description; and the description especially near at hand is 'the Φ -value of p ', with Φ being the agent's internal desirability function (so that for all p and x : if the Φ -value of p is x then the internal desirability of p for s is x). But even such terms, strictly speaking, probably do not refer to numbers but only to very vague ideas of quantities or intervals of quantities. However, in this case, secondly, there is at least a *quantitative intention* of the agent, which surely he cannot fill. But perhaps *we* can fill it, by developing means for measuring the Φ -values; these Φ -values then would be the touched up interpretation of the agent's internal desirability function.

AQPB7 - AQPB9: Determination, quantitativity of the desirability function, and splitting condition: *The desirability functions taken as preferential basis for the intrinsic desirability must be of such condition that there is an interpretation of them not garbling their underlying intention and fulfilling the following conditions: 1. (AQPB7:) The reconstructed function is quantitative. 2. (AQPB8:) It is determined; i. e. the objects for which the function is defined are determined in such way that it is defined if two of them overlap or not, and what the overlapping or not overlapping parts consist in. 3. (AQPB9:) The function fulfills the splitting condition; i. e. if the 'internal desirability' is defined for two overlapping objects p and q under the circumstances c and d , and if $p \& q \& c \& d$ is possible then it is defined for the overlapping part, and the one or two not overlapping parts too, under circumstances implied by $c \& d$.*

5. Conditions of Adequacy, Group 3: Coherence of the Preferential Basis

The conditions of adequacy now to be considered are counterparts to certain conditions of adequacy for the definition of the 'intrinsic desirability': Those in each case are weaker versions of these, and they are formulated for the internal desirability functions to be taken as the preferential basis - as is already the case with AQPB7 to AQPB9. Otherwise it would be impossible to fulfill together the respective conditions of adequacy for the definition of the 'intrinsic desirability' and AQPB2, that the intrinsic desirability must be a positive-linear transformation of the touched up internal desirability functions. The general aim of these conditions is to guarantee consistency; especially, it shall be ensured that for each pair of an object and a circumstance, only one value of the internal or intrinsic desirability is defined.

5.1. AQPB10, AQPB11: Uniqueness and Additive Coherence of the Preferential Basis

The functions of the intrinsic desirability must be unique in the sense that for each triple, consisting of an object and a circumstance and a subject, only *one* value for the intrinsic desirability is defined. And these functions must be *additively coherent*: If the 'intrinsic desirability' is defined for some objects p_1 to p_n under the circumstance c_1 to c_n , and for

$p_1 \& p_2 \& \dots \& p_n$ under a circumstance d , and the p_i ($1 \leq i \leq n$) do not in pairs overlap and $p_1 \& p_2 \& \dots \& p_n \& c_1 \& c_2 \& \dots \& c_n \& d$ is possible then the intrinsic desirability of $p_1 \& p_2 \& \dots \& p_n$ under the circumstance d equals the sum of the intrinsic desirabilities of the p_i under their respective circumstances c_i ($U_{in} p_1, c_1, s + U_{in} p_2, c_2, s + \dots + U_{in} p_n, c_n, s = U_{in} p_1 \& p_2 \& \dots \& p_n, d, s$).

Additive coherence is required if the sum of the intrinsic desirabilities of all the objects accompanying e. g. an action shall be independent of mere differences in the *description* of these objects, especially if the sum shall be identical regardless whether $p \& q$ is described as one or as two objects.

Against the additivity one might object that, according to this condition, *organic goods* would be excluded as intrinsic goods. What one wants is this: There are two non-overlapping states of affairs p and q with the intrinsic desirabilities for a person s under a circumstance c e. g. of 2 and 1 ($U_{in} p, c, s=2; U_{in} q, c, s=1$); but the intrinsic desirability of $p \& q$ doesn't equal the sum of both, e. g. it equals 5 ($U_{in} p \& q, c, s=5$), which would be forbidden by the additivity condition. The reply to this objection is not difficult: Organic goods are not excluded by the additivity condition. The intrinsic desirability function is misconstrued; one could introduce the not realized states of affairs into the circumstances like this: $U_{in} p, c \& \neg q, s=2; U_{in} q, c \& \neg p, s=1; (U_{in} p \& q, c, s=5$ remains as it was); then the two objects and the three different circumstances together are no longer possible ($\neg M p \& q \& (c \& \neg q) \& (c \& \neg p) \& c$), so this is no longer a case in which the additivity condition must be fulfilled. Additivity is required only for the axiological addition of goods and circumstances which can occur together (in one possible world).

Both conditions, uniqueness and additivity, in a weaker form are also required for the preferential basis of intrinsic desirabilities:

AQPB10, AQPB11: Uniqueness and additive coherence of the preferential basis: *The desirability functions taken as preferential basis for the intrinsic desirability must be such that there is an interpretation of them not garbling their underlying intention and fulfilling the following conditions: 1. (AQPB10:) The reconstructed function is unique; 2. (AQPB11:) it is additively coherent.*

5.2. AQPB12: Closed Circumstances

If the concept of intrinsic desirability contains a variable for circumstances, then the possible circumstances should not lead to incoherences. I. e. if the 'intrinsic desirability' is defined for an object p under the circumstance c and for p under the circumstance d and $p \& c \& d$ is possible then the intrinsic desirabilities must be the same (= *condition of closed circumstances*). The condition of closed circumstances especially must be satisfied if d , the second circumstance, is a consistent enlargement of c (e. g. with $d=c \& e$); i. e. the information c was already sufficiently rich or closed so that its enlargement to d could not alter the intrinsic desirability. Otherwise p under the circumstance $c \& e$ would have *two* intrinsic

desirabilities: that of p under the circumstance $c \& e$ and, because $c \& e$ entails c , the differing desirability of p under the circumstance c . So if the two desirabilities really *shall* differ, their appertaining circumstances must be widened so that they are incompatible, e. g. c could be widened to $c \& \neg e$.

The condition of closed circumstances also rules out that intrinsic desirabilities depend on the time: c could be an event at a given time t_1 and d an event at the time t_2 , so that the closedness of the circumstances implicitly would require that the intrinsic desirability does not change from t_1 to t_2 . Otherwise it might occur that at t_1 it would be totally good to do something to bring about p and that at t_2 it would be totally good to try to prevent p , so that one's actions in a certain sense would fight against one another. And this obviously is far from being rational.

So as not to lead to corresponding problems, the *internal* desirability functions taken as the basis for the intrinsic desirability must be such that their circumstances in a weaker form are closed too. According to this condition of adequacy very important internal desirability functions will be ruled out as inconsistent, namely those underlying affective acts (i. e. acts committed in the heat of passion): Being furious at his son for his being cruel ($=c$) a father might value as internally positive ($=u_p$) that the son must suffer some sort of cruelty himself ($=p$) and therefore, in the heat of passion, slaps his son's face; afterwards, realizing what he has done, the father deplores his act ($=d$) and values as internally negative ($=u_n$) that his son suffered such cruelty ($=p$). This constellation may be adequately described as follows: The internal desirability of p under the circumstance c equals u_p , whereas under the circumstance d it equals u_n ; but c and d are compatible; therefore the internal desirability function containing these desirability values has no closed circumstances, and hence is incoherent. And there is no way of closing the circumstances without garbling the underlying intention: E. g. changing the desirability functions in this way: the internal desirability of p under the circumstance $c \& \neg d$ equals u_p , under the circumstance $\neg c \& d$ it equals u_n - this would not help, because we would not know what the internal desirability of p under the circumstance $c \& d$ would be. The internal desirability functions underlying affective acts remain hopelessly unclosed and incoherent, and therefore may not be the basis of rational decision. And I think this incoherence is one deeper reason why often we consider affective acts as irrational.

The condition of closed circumstances may perhaps seem to be too technical and therefore too strong to be applied to internal desirability functions. But I think this is not true; this condition of adequacy can also be understood as an *invitation to clarify* what circumstances are important for one's internal valuations.

AQPB12: Closed circumstances: *The desirability functions taken as preferential basis for the intrinsic desirability must be such that there is an interpretation of them not garbling their underlying intention and fulfilling the condition of closed circumstances: If the 'internal desirability' is defined for an object p under the circumstance c , and for p under the*

circumstance d , and p & d is possible, then the internal desirabilities must be the same ($U_{inap,c,s} = U_{inap,d,s}$).

6. The Resulting Preferential Basis

I do not see any means for proving that these conditions of adequacy are complete; so the question of whether they really are must be open to discussion. Alternatively one might suspect that these conditions are already too strong to allow just *one* desirability function. Of course, in the end this question must be decided by careful *empirical research*. Therefore, I can reply to the objection only with some speculation, based on my personal experience and reflection, which might also, naturally, be open to further changes. According to this experience and reflection, exactly one desirability function would be singled out as the preferential basis of the intrinsic desirability, namely a *corrected hedonistic* one: Under normal circumstances the value of this function with respect to certain feelings is something like the integral with respect to the (positively or negatively) directed intensity of that feeling over the time; but if the components of the personality responsible for my feelings are "restricted" in some way, those values decrease. So my intrinsic desirability would be a 3-adic function with 1. my feelings as the only objects of internal respectively intrinsic desirability (therefore it is hedonistic), 2. the ways of their genesis as the circumstances (they may correct the prima-facie hedonistic desirability), and 3. myself as the subject of desire. (The intrinsic desirability according to a *simple* hedonism, instead, in principle would be a *binary* function - without the circumstances.) However, to confirm such speculations we need a very strong empirical theory of internal valuations, which we do not actually have (cf. e. g. Heckhausen (1989), pp. 455-466).

References

- Fishburn (1988). Fishburn, Peter C.: *Nonlinear Preference and Utility Theory*, Brighton 1988.
- Gärdenfors/Sahlin (1988). Gärdenfors, Peter, & Sahlin, Nils-Eric (eds.): *Decision, probability, and utility*, Cambridge [etc.] 1988.
- Heckhausen (1989). Heckhausen, Heinz: *Motivation und Handeln*, (1st ed. 1980), 2nd, completely revised and enlarged ed. Berlin [etc.] 1989.
- Jeffrey (1983). Jeffrey, Richard C.: *The Logic of Decision*, (1st ed. 1965) Chicago, & London 21983.
- Kusser (1989). Kusser, Anna: *Dimensionen der Kritik von Wünschen*, Frankfurt/Main 1989.
- Raiffa (1968). Raiffa, Howard: *Decision Analysis*, Reading/ Mass. 1968.
- Savage (1972). Savage, Leonhard J.: *The Foundations of Statistics*, (1st ed. 1954), New York 21972.