Intentions Are Optimality Beliefs - but Optimizing what?

Christoph Lumer

(Appeared in: Erkenntnis 62 (2005). Pp. 235-262.)

Abstract: In this paper an empirical theory about the nature of intention is sketched. After stressing the necessity of reckoning with intentions in philosophy of action a strategy for deciding empirically between competing theories of intention is exposed and applied for criticizing various philosophical theories of intention, among others that of Bratman. The hypothesis that intentions are optimality beliefs is defended on the basis of empirical decision theory. Present empirical decision theory however does not provide an empirically satisfying elaboration of the desirability concepts used in these optimality beliefs. Based on process theories of deliberation two hypotheses for filling this gap are developed.

1. Criticisms of the Desire-Belief Model of Action, the Proliferation of Theories of Intention and the Scope of this Paper

In philosophy during roughly the last 15 years the desire-belief model of action (fostered e.g. by Davidson (1963, in particular 3-12; 1978, 87), Goldman (1970, 49-57; 72), Churchland (1970), Brandt (1979, 47 f.; 64-66), Audi (1986, 98) and Lennon (1990, 37-39)) has been heavily attacked, in particular because it tries to do without intentions, thus underrating their role which cannot be filled by sets of desire-belief pairs: Desire-belief pairs do not have the necessary resoluteness, the action chosen by them cannot be executed later on, they do not explain difficult decision processes etc.

During the same period a wealth of philosophical theories about what type of propositional attitudes intentions are has been developed. Some main answers to this question are: *Sui-generis theory:* Intentions are propositional attitudes in their own right, not reducible to other types of propositional attitudes (Bratman 1987, 10; 20; 110; Donagan 1987, 41; 81; Mele 1992, 127; 162); they consist of cognitive and conative components (Brand 1984, 47; 153; 239; 237; 240 f.; 266-268);[1] *prevision theory:* intentions are - self-fulfilling - previsions of one's own actions (Harman 1976; Velleman 1989a, 109 and 109-142; 1989b).[2] And some older theories of action may be interpreted as theories about the empirical nature of intention, too, e.g. models of practical inference (Aristotle, NE 1111b-1113b; 1139a; 1147a; von Wright 1963; 1971, 96-107; 1972),

[1] Brand's explanations of the conative components of intentions are not very clear; what he writes tends towards a need theory of intention or a drive theory of action (Brand 1984, 266-268). The former will be discussed below, and the latter is a theory which again does not consider intentions. A further problem of Brand's theory is that it is not clear how the cognitive and the conative component are brought together. In Brand's exposition, these components seem to be represented by two different and unconnected mental attitudes. - Because of these problems Brand's theory here will not be discussed further.

[2] Audi earlier defended a combination of the prevision theory of intention and a desire-model of action (Audi 1973, 65).

psychological normativism, which holds that an action is caused by the belief that this action is socially required in the particular situation (Mead 1934, 152-164; a partial psychological normativism is included in Habermas' concepts of 'communicative action' and 'normatively regulated action', cf.: Habermas 1981, 127 f.; 132-134; 143; 148-151; 385-387; 412; 418; Habermas 1975, 280-282), or models of needs, presuming that we act out of our strongest or most developed need (Maslow 1954, ch. 4-7; Kambartel 1974, 62-65).

The present paper is a contribution to the discussion about the right theory of intention. Apart from criticizing some presently discussed theories of intention, it mainly has constructive aims: A strategy for empirically deciding between such theories will be sketched, and a new proposal for an empirical theory of intention will be elaborated and defended as meeting the criteria of that strategy. The theory proposed here is the optimality-belief theory.

2. A Strategy for Empirically Determining the Right Theory of Intention

The proliferation of theories of intention in philosophy was and is accompanied by a rather ramified and in part very sophisticated discussion about the validity of these theories. One particular but very fundamental problem of this discussion is that analytical and empirical questions are not clearly distinguished. We find arguments appealing to the common meaning of "intention" and the conceptually fixed function of intentions next to arguments about the biological or practical function, the predecessors, the accompanying mental states, possible or factual content etc. of intentions. And for many philosophical theories of intention it is not clear if they want to define 'intention' or if they want to make empirical hypotheses about the nature of intentions.

Following an idea of Myles Brand (Brand 1979; 1984, 33; 35 f.), it is proposed that this problem can be solved by the following strategy. In a first step a clear definition of 'intention' is developed covering much of our intuitions about intentions and some empirical information about how intentions work but which is functional, rather formal and empirically open to the question as to which mental states are intentions or function as such. The second step consists in gathering such empirical information about the intentions just defined which may be used to decide between different hypotheses about what propositional attitudes are intentions. So this empirical information would function as adequacy conditions for empirical theories of intention. The final step would be to filter out empirically inadequate theories of intention with the help of these adequacy conditions. - In what follows I will pursue this strategy, which is appropriate for deciding between the main theories of intention. For refining the remaining theory and for deciding between its variants more specific empirical information has to be considered, which will be done in the second half of the paper.

The idea of defining 'intention' in the way just explained is Brand's. His definition itself (Brand 1984, 174), though, I find much wanting in particular with respect to the proposed solution to the problem of deviant realizations of intentions. Therefore, I will sketch my own definition of 'intention'. In this definition the expression "action-generating mechanism" is used, which can roughly be explained as follows:

An *action-generating mechanism* is

1. a system with a mental and an executionary subsystem;
2. the executionary subsystem can internally cause a certain range of types of behaviour;
3. the mental subsystem is capable of attitudes towards the behaviour of the executionary subsystem;
4. and both subsystems together form a cybernetic system in which mental attitudes of a certain type Φ towards behaviour of the executive subsystem - in a certain range of types of behaviour - rather reliably and internally cause this behaviour in such a way that the attitude provides the set point for the executive subsystem.

Intentions then (by definition) are mental attitudes of the type Φ in an action-generating mechanism towards some behaviour of the mechanism. So intentions (analytically) are attitudes that in principle are able to provoke the respective behaviour. Defining 'intention' is not my present concern, and the details of the definition just given are not important for the following discussion. Therefore, I will not defend it here (but cf. Lumer forthcoming, ch. 4.3). But the empirical part of the question 'what are intentions?', according to the definition just given, now can be specified like this: what type Φ of mental attitudes empirically are intentions (in the sense just defined)? Or what type Φ of mental attitudes in human beings empirically fulfils the analytically specified function of intentions?

The following empirical properties of human intentions can be used as conditions of adequacy for answering this question. For reasons of simplicity, I only mention the empirical properties; the adequacy conditions, more precisely, then should say that mental attitudes of the proposed type must have this property and that it must be easily (i.e. with very probable assumptions) explainable how they can have it.

AQ1: Present and distal intentions: Humans have present as well as future intentions.

AQ2: Various logical forms of intentions: Human intentions have several logical forms. Apart from simple intentions, there are e.g. general, conditional and disjunctive intentions.

AQ3: Transition deliberation to intention: Often deliberation precedes an intention in such a way that there is at least a seamless transition between them. The intention may even be equal to the cognitive result of the deliberation. And new deliberation can dissolve old intentions as such.

AQ4: Settling effect of intentions: Intentions to a certain degree settle what will be done. If there has been a consideration of alternatives they (mostly) finish it and include a commitment to the intended action. (Bratman 1987, 4; 16 f.)

AQ5: Degrees of intentions: Human intentions come in different degrees of firmness. (AQ5 by the way specifies the restriction of AQ4 that settling what will be done takes place only "to a certain degree" etc.)

AQ6: Intentions as mental dispositions: Forming an intention is a mental event; but after that the intention is not an experienced mental state but a disposition which may direct behaviour unconsciously.[3]

AQ7: Small and big actions: Human intentions aim at a very wide spectrum of intended actions, from scratching oneself or jumping over a brook up to writing a book and doing what one can for diminishing hunger in the world.[4] And often small actions are intended as parts of bigger actions.

AQ8: The weighing content of deliberation: One important content of deliberations preceding intentions is to ponder possible advantages and disadvantages of the considered options.

Intentions are always about actions. But these actions can be described in different ways. The two extremes are descriptions via an end, e.g. 'decreasing hunger in the world' (which means: doing *something* which has the effect that hunger in the world is diminished), and providing parameter specifications that are understandable to the executionary subsystem and thus can be executed immediately, e.g. 'signing this cheque using the pen in front of me'. Gollwitzer has called these types of intentions "*goal intentions*" and "*implementation intentions*" respectively (Gollwitzer 1999, 494 f.).

AQ9: Uniqueness of intended actions: In their implementation intentions, as opposed to goal intentions, subjects try to make exactly one of the possible alternatives standing out so that the executionary system can execute it (this implies that formerly disjunctive goal intentions (cf. AQ2) have to be restricted to one option); otherwise there would not be a (real or effective) decision or directive to the executionary system and the latter would have to "decide" on its own.[5] [6]

[3] Forming an intention may have additional mental effects like getting calm - as they are highlighted in the Rubicon-theory of Gollwitzer and Heckhausen (Gollwitzer 1991; 1996; Heckhausen 1989, 203-218; Heckhausen / Gollwitzer / Weinert 1987). But these effects are not identical to the intention.

[4] The "big" actions do not have to be specified in the form of conjunctions of "smaller" ones; on the contrary, at least initially they are mostly specified only by their aim, e.g. as an 'action which will lead to aim *x*'. Therefore, AQ7 is independent of AQ2.

[5] Condition AQ9 does not deny that it is *possible* to have implementation intentions to execute two or more mutually exclusive alternatives. But cases of this sort are *action slips*, which we try to avoid. Such action slips may e.g. occur because the action's description is not sufficiently specific and, therefore, ambiguous. This may be the case when someone - absent-mindedly - intends to 'replace this lid on the container' and as a consequence puts the lid of the sugar container on the coffee cup (where the two containers are similarly shaped) (Norman 1981, 7; 8).

[6] Condition AQ9 does not hold for goal intentions. It may be perfectly rational to have two goal intentions for mutually excluding actions (with perhaps, additionaly, mutually excluding aims), namely if we do not know which implementation intentions would be the correct specifications of the goal intentions. Think of a person who wants to enter a building with a double-winged door. The person knows that the two wings open in the same direction but she does not know in which direction. For accelerating things she pushes the right wing with the right hand and at the same time pulls the other wing with the left hand, exactly knowing that if she succeeds in pushing pulling will be unsuccessful and vice versa. The two actions described in the goal intentions, namely 'to push the right wing open' and 'to pull the left wing open' are (not logically but empirically) mutually excluding, whereas the actions described in the implementation intentions, namely 'to press with the right hand the right wing's surface' and 'to grasp the handle of the left wing with the left hand

The properties just mentioned are not all empirically known properties of intentions but those that may suffice for deciding empirically between the major theories of intention.

3. Criticisms of Some Philosophical Theories of Intention

Philosophical theories of intention come in two big groups, doxastic and non-doxastic theories. *Doxastic theories of intention* hold that intentions are certain types of beliefs that the intended action (i.e. the state of affairs that the respective agent will perform an action of type *A* at time *t*) has a certain property *F*. They differ in their respective assumptions about that *F*. The *prevision theory* holds that *F* is empty or equal to 'is a fact' (the intention then is the belief 'I will do *A* at *t*' or 'It is a fact that I will do *A* at *t*', respectively, so that the beliefs' contents are analytically equivalent); *psychological normativism* holds that *F* is equal to 'is socially required or prescribed' etc.

and push', obviously, are not. (Cf. Ginet 1990.) One of the actions described in the goal intentions actually is even impossible but the subject does not know which one. And, as a consequence, the subject does not know which implementation intention is a specification of the realizable goal intention. Expressed differently: Both implementation intentions are realizable; but the realization of one of them will not have the desired success; and the subject does not know which one. Therefore, it is completely rational to execute both implementation intentions for increasing chances to realize one of the mutually excluding goals, thereby having two "inconsistent" goal intentions.

Bratman takes it to be a rationality requirement that intentions have to be strongly consistent, relative to the subject's beliefs, i.e. it should be possible to execute all the intentions in a world in which the subject's beliefs are true (Bratman 1987, 113). He illustrates this requirement with an example which is analogous to the example of the double-winged door (ibid. 113 f.). - Bratman's claim is reasonable for *implementation* intentions only because the executionary system otherwise would not "know" what to execute and thus might execute no alternative at all or a bad one, whereas the claim is not reasonable for *goal* intentions because a rationality requirement of this sort would impede rising our chances to realize at least one of some mutually excluding goals and goal intentions. Bratman's argument for his stronger claim is that without it intentions would not be suited for coordinating plans (ibid. 113). But a complete coordination is possible only if we know in which possible world we live - which is not the case. Subjects with incomplete knowledge have to content with incomplete coordination. If we have to reckon with several possible worlds rationality and coordination require only that everything intended *for each of the possible worlds* is consistent (otherwise it would not be a possible world). But they do not require that everything intended *at all* - plus the possible worlds in which these intentions may be fulfilled - must be consistent (if we have to deal with at least two possible worlds they cannot be consistent). The technical problem is that Bratman does not differentiate between goal and implementation intentions and that his rationality condition neither reckons with merely probabilistic beliefs nor with intentions for different possible worlds. - Bratman uses the rationality requirement defended by him for arguing that behaviour in cases like the one explained is rational and intentional but that there is no corresponding (goal) intention (ibid. 113-119). Firstly, I have given reasons why Bratman's rationality requirement is false so that the premise of his reasoning is false. Secondly, in the example discussed there are some rather obvious goal intentions, namely 'to push the right wing open' etc. (cf. above). This means that Bratman's further claim about the missing (goal) intention is false, too, for independent reasons.

Doxastic theories of intention comply easily with the first seven adequacy conditions: AQ1 (Present and distal intentions): Present and future intentions, according to doxastic theories of intention, differ in that the predicate '*F*' is applied to the description of an immediately following or a future (at time t_f) action of the subject. And future intentions are executed when the subject comes to know that 'now' is t_f. AQ2 (Various logical forms of intentions): Conditional intentions e.g. are explained as being beliefs with the content: 'If condition *c* holds then the action *a* is *F*.' If the subject then comes to believe that *c* holds, according to logical rules, he is justified to believe that *a* is *F*; and often he will execute that inference thus arriving at an unconditional intention. Intentions of the other logical forms can be explained analogously. AQ3 (Transition deliberation to intention): The seamless transition from deliberation to the intention, according to doxastic theories, is due to the fact that the intention is identical with the result of the deliberation where the deliberation consists in investigating the truth of the proposition '*a* is *F*'. AQ4 (Settling effect of intentions): Forming an intention can stop the consideration of alternatives because after pondering various pros and cons the subject has finally come to believe that *a* is *F*. And this belief, according to the subject, may be more or less justified and thus stable with respect to reconsideration - unless important new information is acquired. AQ5 (Degrees of intentions): On the other hand evidence may be rather weak so that the subject after coming to believe that *a* is *F* is uncertain about this result, which, according to the doxastic theory, is identical to have a less firm intention. AQ6 (Intentions as mental dispositions): Forming a belief is an act that may trigger the corresponding action. But the resulting belief is a disposition that may be the structuring cause of the corresponding action, which is triggered by a different stimulus, e.g. becoming aware that some condition is fulfilled. AQ7 (Small and big actions): Predicate '*F*' may apply to small as well as big actions. And there are several important candidates for '*F*' for which holds: If *F* applies to a big action *F* applies to all its (relevant) parts, too, so that the coming into existence of the small intentions can follow the paths of logic.

The doxastic theories of intention mentioned so far, i.e. the prevision theory, psychological normativism, models of practical inference and models of needs, however do not comply with adequacy condition AQ8.[7] According to none of them, enquiring the truth of the proposition in question (that *a* has the property *F*) would require pondering the advantages and disadvantages of the various options: Normal prevision requires an inference from beliefs about circumstances and general (statistical) laws, self-fulfilling predictive judgements, however, could be completely arbitrary, thus expressing freedom of the will and making our body follow this *liberum arbitrium*; judgements about the social requirement of an action mostly are deontic conclusions from beliefs about some social standard and the given situation; and practical inferences (of the standard form) as well as beliefs about need satisfaction, finally, require deduction from the aim or need and a proposition that a certain action would fulfil them. Psychological normativism, models of practical inference and models of needs, in addition, cannot explain how we choose exactly *one* action as the

[7] The aim of this paper ist primarily constructive. The following criticisms, therefore, are only sketchy, focus on few essential points only and cannot do complete justice to the theories discussed.

one to be executed (i.e. they violate AQ9). All the restricting mechanisms appealed to by these three theories, i.e. social norms, aims and needs, fortunately, leave much room for various alternatives.

The currently most famous non-doxastic theory of intention is the *sui-generis model*, holding that intentions are irreducible propositional attitudes of their own (Bratman, Mele, Donagan). Apart from the fact that introducing a new propositional attitude seems to be a bit *ad hoc*, such theories have some difficulties to comply with the first three adequacy conditions and the adequacy conditions AQ7 and AQ8. These difficulties are not insurmountable in all cases but every condition requires an extra explanation and an additional part of the theory - which until today has not been provided. The *sui-generis* theory e.g. must hold that conditional intentions (cf. AQ2) are attitudes of type Φ with the propositional content 'if p then I will do A at t'; and it must give an explanation why after coming to believe that p the subject develops an attitude of type Φ with the propositional content 'I will do A at t'. There may be an explanation for this but it is one that has to be explicitly developed. Analogous difficulties arise with the conditions AQ1 (present and distal intentions) and AQ7 (small and big actions). But above all, the *sui-generis* model of intention does extremely badly with the third adequacy condition about the seamless transition from deliberation to intention and, consequently, with AQ8 (the weighing content of intentions). According to the *sui-generis* model, the intention is quite isolated from the deliberation. This problem may be diminished e.g. by complementing the model with a decision-theoretic model of decision-making where desire-belief pairs are weighed against each other. If the result of the decision process is not the action itself, i.e. if there is the intermediate step of forming an intention, this result could be a value judgement that a certain alternative has the highest subjective expected utility (or something similar). But then this value judgement could function as an intention and the *sui-generis* intention would be superfluous.

4. Empirical Theory of Decision - 1. The Way to an Adequate Theory of Intention

The empirical theories of intention scrutinized so far have all been proposed by philosophers. They are empirical in the sense that they consist of empirical hypotheses but not in the sense that they rely on genuine empirical research. None of these theories was empirically adequate. Therefore, it is time to see what empirical scientists have found out. Empirical decision research in psychology and economics has provided wealth of data and theories about which decisions are taken on the basis of given desires and beliefs (overview: Camerer 1995, 617-674; Slovic / Lichtenstein / Fischhoff 1988). The mainstream opinion is that decisions consist in optimizing, i.e. maximizing some sort of - expressed neutrally - desirability.[8] The concepts of

[8] "*Optimizing*" here shall mean: maximizing desirability; and "*maximizing*" without further qualifications here is understood as *weak maximizing*, i.e. choosing an option with a value at least as high as that of every alternative, which allows for having more than one object with maximum value. (Sen distinguishes between

'desirability' vary from theory to theory, ranging from traditional subjective expected utility to e.g. the prospects in Kahneman's and Tversky's prospect theory (Kahneman / Tversky 1979; Tversky / Kahneman 1992), which is the most famous and seemingly most promising non-linear decision theory (i.e. a theory which holds that probabilistic outcomes of an action are *not* weighted proportionally to the probability of these outcomes).

One important rival to optimizing is satisficing (advocated e.g. by Simon (1956); Slote (1989) proposes satisficing as a normative ideal). This theory says that people do not strive for optimum alternatives but that they stop deliberation, decide and begin execution when they think they have found an action fulfilling their requirements in a satisfying way, i.e. an action coming up to a certain aspiration level. - At least one half of this theory is true but in particular its opposition to optimizing partly rests on a confusion. Who would choose a certain course of action if he believes a given alternative to be better?

The seeming contradiction can be disentangled if various meanings of "optimize" are first differentiated. The *general sense* of this expression is: to do something (to develop alternatives, deliberate etc. and decide) with the aim of getting the optimum result. The sort of optimizing which we would like to do is to simply choose the really best option of all options available; this may be called "*genuine (or objective) optimizing*". Unfortunately, genuine optimizing normally is not possible because the required information is not simply at hand. A first answer to this impossibility is *straightforward optimizing*, i.e.: to investigate as long as one is (relatively) sure to have found the really best solution. Straightforward optimizing in most situations of daily life is not possible either because the necessary information is not obtainable or because we have to act earlier. And if it is possible it is usually not very efficient. A second and often more efficient answer to the mentioned impossibility then is *basic (or subjective) optimizing*, i.e. in cases where simply choosing the really best alternative is impossible to use a substitute for this criterion which shall lead as near as possible to the best alternative, namely: to choose that alternative which, on the basis of the

"optimizing", which he uses in the same sense as is done here, and "maximizing (desirability)", which better should be named: "*negative maximizing* (desirability)", i.e. choosing an option for which no better alternative exists (Sen 1997, 172; 182). Sen advocates negative maximizing (instead of (positive) maximizing) because in cases of incomplete preference orders negative maximizing would prevent a deadlock. This proposal may have some formal advantage but it is neither convincing from an empirical, action-theoretic nor from a rational point of view. In forced choices people manage to overcome their reluctance towards ranking two options. Usually they finally rank them relying on some idea of the options' desirabilities (e.g. the interval in which the real desirability may lie). They are reluctant to rank in this way because it is very inaccurate; but because of time pressure they are forced to do so. The empirical hypothesis then is that in real decisions finally and strongly incomplete preference orders, which would impede (positive) optimizing, do not exist. From a rational point of view negative maximizing seems to be based on a fallacious *argumentum ad ignorantiam*. ('Since I do not know a to be worse than b, therefore a is at least as good as b' (cf. Sen 1997, 185); one of the problems of *argumenta ad ignorantiam* is that they can prove even the exact opposite: 'Since I do not know that a is at least as good as b, therefore a is worse than b.') And negative maximizing would be a bad advice e.g. in cases where there is a normal preference order in $\{a_1, ..., a_n\}$ with a_1 being a good and the best option and where there is some b preferentially completely disconnected from a_1 to a_n, so that negative maximizing would regard b as being on a par with a_1 - which is rather far-fetched, though.)

available information, has or is believed to have the highest desirability prospects among the options considered - where the '*desirability prospects*' may be conceptualised as expected desirability or as some other, non-linear derivate of the real desirability which can be determined on the basis of weak information. Relativity to the available information and the concept of 'desirability prospects' here shall imply that in (extreme) cases where all the necessary information is available basic optimizing coincides with genuine optimizing. - Basic optimizing may be undertaken without any deliberation or after deliberation of various lengths and qualities. As a consequence the results, namely the options finally chosen, may be the better the better and longer the deliberation was (but probably with diminishing returns). On the other hand at least longer deliberations are more costly. A first solution to this problem of the diverging consequences of deliberation is *satisficing*, i.e. to stop deliberation when a certain aspiration level has been reached. A second and probably better solution is to change the scope of optimizing, namely the sort of entities that are to be optimized: 1. The original set of alternatives consists of executive actions alone, 2. the more sophisticated one consists of totals of executive actions plus the eventually still necessary continuation of the deliberation to find out further details. (The latter set also includes options for which no continuation of the deliberation is necessary.) Applying basic optimizing to the second set of alternatives may be called "*reflexive optimizing*",[9] which means: to deliberate so long until the total consisting of the executive action and the deliberation has or is believed to have the highest desirability prospects. This is identical to optimizing the deliberation proper, i.e. to deliberate and decide in that way (so long and with such a quality) that has the highest desirability prospects, because the executive action (with its consequences) is included in the set of the deliberation's consequences, which have to be considered in estimating the deliberation's desirability.[10] [11] - This differentiation of various variants of optimizing reveals a deep ambiguity in the characterization of the mainstream of empirical decision theory as "optimizing" theories. What the advocates of these theories usually mean however is basic optimizing (which includes reflexive optimizing) and not straightforward (or genuine) optimizing.

Now let us try to disentangle the seeming contradiction between optimizing and satisficing. Straightforward optimizing is possible and is really opposed to satisficing. But as just mentioned

[9] A more detailed exposition of reflexive optimizing, where reflexive optimizing is equated with practical rationality, can be found in: Lumer 1990, 390-399.

[10] Some people doubt that reflexive optimizing is possible because they think the required information is not available (cf. e.g. Gigerenzer et al. 1999, 10-12). Indeed, the subject can never be sure to have found the really best deliberation (or total consisting of the executive action plus deliberation). But the subjects can use rules of thumb for determining which extent and quality of deliberation (plus the decision according to the information obtained by this deliberation) may be really optimum. And these rules of thumb may be justified by estimating desirability prospects for them on the basis of experiences with deliberations undertaken so far.

[11] One might think of another logically possible combination, namely to apply something like straightforward optimizing to the total consisting of executive action and deliberation, too. But this is nearly impossible empirically because of the reflexive nature of this kind of optimizing. For knowing what the really best of such totals is much, much more information would be necessary than this best alternative would permit to acquire. Only if we could acquire this information for free or nearly free such kind of optimizing would be possible.

the mainstream of empirical decision theory does not hold that people generally optimize straightforwardly but that they optimize in the basic way. Only some neurotic perfectionists try to optimize straightforwardly - falsely disregarding the deliberation costs. Basic optimizing however does not imply straightforward optimizing and is compatible with satisficing because it is possible, often done and often rational at the same time:

b1: to believe that, according to the available information, an action *a* is optimum in the sense of having the highest desirability prospects among the options considered (basic optimizing);

b2: to believe that *a* probably is not *really* optimum, i.e. of the highest desirability, among the options available (negative part of satisficing); and

b3: to know or believe that *a* really is at least quite good among the options available (satisficing).

States of affair *b1* and *b2* are compossible only if, as is normally the case,

b4: the subject does not know which alternative is the really best among the options available (no genuine optimizing) and, therefore, decides according to the desirability prospects among the alternatives considered (switch to basic optimizing).

Finally, reflexive optimizing is not only compatible with satisficing but also the best way to satisfice and the rational way to take decisions. This holds because satisficing is already an endeavour to solve the efficiency problem of deliberation and because reflexive optimizing optimizes the deliberation, too, i.e. it provides the best solution at hand to this problem.[12] - In the case of reflexive optimizing subjects think about the desirability of their deliberations and at specific times they may

b5: believe that deciding right now and executing the alternative which among the options considered has the, according to the information available at the present moment, highest desirability prospects (cf. *b1*), has a higher desirability prospect than continuing the deliberation - and therefore decide immediately.

b5 is only a specification of *b1* with respect to deliberations. And the subject again may believe *b1* to *b4* about the executive action chosen via reflexive optimizing. In particular he may believe that most of the executive actions chosen with the help of this strategy are not really optimum (cf. above, *b2*) but only satisfying (cf. *b3*) though he does not know *which* alternative is better (cf. *b4*). But in addition he often will be justified in believing that finding a really better executive action via continuation of the deliberation will have higher costs than the gains reached by such an improvement (cf. *b5*). Therefore, clever people try to optimize reflexively and not so clever people usually at least consider the extent of their deliberations with an intuitive understanding what the costs and benefits of extended deliberation are - thus at least approaching a little way towards reflexive optimizing.

So theories of basic optimizing seem to be a good approach for determining what deliberation is striving at. But most of the present empirical theories of optimizing are *hydraulic* in

[12] Gigerenzer et al. (1999, ch. 13) provide an example of a sophisticated form of satisficing, where the aspiration level is fixed in such a way as to optimize the whole procedure of decision (though still many components of the decision costs are ignored). Reflexive optimizers then may use the heuristic developed there (or an improved version of it).

the sense that they presuppose that the decision is taken unconsciously by addition and counterbalancing of motivational forces; they do not say anything about the subjective processes during the decision, in particular they do not mention intentions. They just consider the input of the decision process, namely desires and beliefs, and its output, which is the action. This means the criticisms directed against desire-belief models of action apply here as well.

But there is a straightforward solution to this problem, which is to introduce an appropriate intention. This is done in my first hypothesis:

H1: Optimality hypothesis: Forming of an intention consists in an optimality judgement that a certain action is optimum [13] (in a sense to be specified) among the considered set of open alternatives. And the intention is the resulting optimality belief. But they are intentions only - roughly - until that moment when the action is finished or, if it is never finished, when the prospected end of its execution has been reached or, if the subject believes in deictic form something with respect to these two dates (end of action, projected end of action), when these assumed moments have been reached.

The ultimate condition e.g. implies that if the subject erroneously believes that he has not yet executed the optimum action the intention to execute it still holds; and if he erroneously believes that he has already executed the action the optimality belief ceases to be an intention, i.e. the intention is deactivated, i.e. believed to be fulfilled (for examples cf. Heckhausen 1987, 158 f.; 164-167).

[13] Buridan cases, i.e. situations in which two (or more) options are judged to be weakly optimum so that no strongly optimum option (better than each alternative) exists, require some refinement of the optimality hypothesis (*H1*). There are at least two possibilities: 1. "Optimality" in *H1* is meant in the weak sense; and for fulfilling the uniqueness condition (AQ9), for Buridan cases a proviso has to be introduced, namely that in these cases the final intention is formed by a further mental device, internal to the decision system, that picks one of the weakly optimum actions, independently of desirability considerations, for making it the object of our intention. (This is the proposal of Ullmann-Margalit / Morgenbesser (1977).) This interpretation unpleasantly complicates the empirical hypothesis by introducing a two-stage criterion for intention formation. 2. "Optimality" in *H1* is meant in the strong sense, which implies that in hard Buridan cases (with more than one option being weakly optimum) people are unable to decide directly; if they are not able to dissolve the equivalence by refining the assessment or by narrowing for some while their field of attention to one of the options (thus reducing the set of open alternatives) they have to go back to using some random device external to the decision system, whose usage then would constitute the strongly best option; this random device may be physically external, like tossing, or psychological though external to the deliberation system, e.g. taking always the option that has first appeared to the mind. (This is Rescher's (1959/60) proposal.) This interpretation from a rational point of view seems to be absurd because using an external random device only complicates the action and is not strongly better than executing directly one of the weakly best options. But the problem seems to be that in hard Buridan cases the latter options may no longer be available. - So according to the first interpretation of *H1* the random device is internal to the decision system, whereas according to the second interpretation it is external to the decision system and the decider has to choose to use some such device and to adopt its outcome. Which of these two interpretations of *H1* is right is an empirical question. The problems many people have in Buridan situations and their actually using external (to the decision system) random devices are evidences of the second interpretation.

This proposal has several advantages. Firstly, the optimality theory gives room to intentions thus avoiding the problems of desire-belief theories of action and of hydraulic theories of decision. Secondly, taking optimality *beliefs* to be intentions implies to preserve the initial success of doxastic theories of intention in complying with the first seven adequacy conditions, which partly could not (or only with large difficulties) be satisfied by any non-doxastic theory. Thirdly, relying on empirical research about the content of decision processes, in particular on the decision theoretical idea of pondering advantages and disadvantages of alternatives, and therefore taking *optimality* beliefs as intentions resolves the problem of the competing doxastic theories. (These could not satisfy at least one of the last two adequacy conditions because the content of intentions assumed by these theories was too far away from the empirical deliberation process.) In contrast to these theories, the optimality theory connects the content of intentions with the content of deliberations and, therefore, complies with the last two adequacy conditions, too. In short, the optimality theory is the only empirical theory of intention considered here complying with all the nine adequacy conditions.

But there are some well-known objections to the optimality-belief theory of intention. At least some of them should be discussed here. A first objection says that in hard cases of weakness of the will, particularly when we are under the influence of strong emotions, we believe a certain action to be optimum but intentionally perform a different action, which seems to exclude that the latter action rests on an optimality belief. But this would be a hasty conclusion. The action actually performed may be intended via a different optimality belief relying on different criteria of desirability so that the present theory has to determine which sort of desirability is considered in those optimality beliefs making up intentions. An initial discussion of this issue can be found below (sect. 6 and 7).[14] The same answer may be given to a second objection, namely that we may have rather detached optimality beliefs that do not motivate us: Again there are two sorts of optimality beliefs relying on different desirability criteria. A third objection says that we may reflect about best actions without a view toward intention formation, e.g. in a nonpractical project of self-analysis (McCann 1998, 221). In so far as this objection differs from the second one the case referred to may be explained by the fact that the agent is thinking about merely hypothetical alternatives and not about alternatives which he already believes to be open to him. A fourth objection points to the possible complexity of optimality beliefs: Tiny children, when they are just barely able to intend, may not dispose of the concepts figuring essentially in optimality propositions (Audi, personal communication). The second part of the present theory will solve this problem: in the simplest case the optimality belief will be reduced to believing that an alternative a_i has a certain advantage which remaining in the current state does not have (cf. below, sect. 6). Even tiny

[14] A more satisfactory answer to the problem of weakness of the will, though, presupposes a theory of intrinsic desirability, which cannot be given here. A proposal developed elsewhere (Lumer 1997) says, weakness of the will may arise because strong emotions change our intrinsic desirability judgements during the occurent emotion, or more precisely they induce additional intrinsic desires. As a consequence optimality beliefs relying on stable and long-term intrinsic desirability criteria are disregarded in favour of a short-term optimality belief partially based on the emotion-induced intrinsic desires.

children are able to have such beliefs. In addition, the optimality hypothesis e.g. permits optimality beliefs arrived at by implicit learning, which leads to a vague impression that a certain option is better than its alternative, but where the subject cannot rationally justify this impression (cf. e.g. Bechara et al. 1997; Reber 1989, 229-231).

5. Empirical Theory of Decision - 2. Some Shortcomings

My first hypothesis leaves open what kind of desirability (including desirability prospects) people try to maximize in deliberation and, accordingly, what concept of a 'best' or 'optimum' action people use in their intentions. A straightforward way to fill this gap should be to define 'desirability' according to the respective findings of empirical decision theory about the deliberation process. But this strategy turns out to be unsuccessful.

Firstly, there is an enormous wealth of competing empirical theories of decision that are in the tradition of rational decision theory (overview: Camerer 1995, 617-651). All these theories truly assume that decisions are made by pondering the advantages and disadvantages of the alternatives coming to the subject's mind, that these advantages or disadvantages (mainly) are constituted by the (intrinsic) positive or negative desirability of the possible consequences of the options, and that the subjective probabilities of these consequences influence their weight in the final decision in a roughly monotonous way. But they differ in their assumptions how these things are to be combined; i.e. they differ exactly in their assumptions about what desirability function people try to maximize in their deliberation.

Secondly, because subjective expected utility theory has been falsified too often as an empirical model of human decision (overview: Camerer 1995, 622-626; 644-649; 652-670) a wealth of alternative models, non-linear or generalized expected utility theories, has been developed. Though some of these theories are elegant and many of them mathematically rather ambitious several studies have found out that none of them is prognostically satisfying (cf. e.g. Camerer 1992; Harless / Camerer 1994; Currim / Sarin 1989). In normal cases they are not prognostically better than subjective expected utility theory. This holds even for the most prominent among these theories, i.e. prospect theory.

Thirdly, the theories of decision in the tradition of rational decision theory are not hydraulic by chance. The hydraulic version has been used for coping with the fact that people do not consciously calculate the defined desirability values proposed by the various empirical models of decision. Most people would not even understand the formulas used for description, in particular the formulas proposed by non-linear decision theories. This means that these models do not reflect what goes on in conscious deliberation and which concept of desirability is used in the resulting optimality belief, which in turn is identical to the intention.

6. The Contribution of Process Theories of Deliberation - Deciding How to Decide

The second important stream in present empirical decision theory is process tracing research, which with several techniques tries to find out what people actually think, consider or calculate and which decision strategies they follow during deliberation.[15] Until now *the* one theory of the decision process is not within reach but the mentioned type of research has provided a huge wealth of empirical findings, about very diverging decision strategies. One such strategy e.g. is to consider a particular, the perhaps most important, dimension of the options first, e.g. the costs of a flat to be rented, and to eliminate all the options that do not meet a certain minimum standard or, which is a variant of this strategy, to choose the alternative performing best in this dimension. Another, much more costly strategy is to compare alternatives pairwise trying to establish a dominance or an artificial dominance relation between them - where an artificial dominance relation holds if for every aspect of alternative *a* there is an aspect of alternative *b* which is at best equally good as the related aspect of *a* and where the related aspects are not always identical. The simplest strategy is to choose an option immediately when it comes into mind if it has at least one positive aspect and if no (grossly) negative aspect of it comes into mind either. Each of these strategies implies a (secondary) criterion for the optimum alternative. And the diversity of the strategies entails a similar manifoldness of such secondary criteria of what is best. The strategies differ in applicability, decision expenditures and in their exactness, i.e. the degree of leading the nearest possible to the really best option. People do not only differ intersubjectively and biographically in their way of using them but they use different strategies in different situations or even change their strategy perhaps several times during one deliberation. In addition, people use the different strategies according to the difficulty and the importance of the decision task. In a certain sense they decide how to decide, and there is a certain tendency to thereby optimize the total consisting of the decision process plus the executive action, i.e. a tendency towards reflexive optimizing (cf. e.g. Beach / Mitchell 1978; Johnson 1979; Payne / Bettman / Johnson 1988; 1993; Shugan 1980; Svenson 1979; 1996). Of course, reflexive optimizing cannot be very precise because it includes omitting information. But humans can accumulate experiences (including experiences of other subjects) with the several decision strategies in which type of situation which type of strategy considered ex post and taking all things together provided good results, and they can learn from such experiences.

This means during deliberation about executive actions people use decision strategies implying *many secondary* criteria of the desirability and optimality of alternatives. But there seems to be *one primary*, i.e. fundamental and most exact, criterion of the desirability and optimality of

[15] Important research of this type has been done e.g. by Beach, Bettman, Gigerenzer, Johnson, Mitchell, Payne, Rubinstein, Shugan, Svenson, Todd, Tversky or Peter Wright (Beach / Mitchell 1978; Gigerenzer et al. 1999; Johnson 1979; Payne / Bettman / Johnson 1988; 1993; Rubinstein 1988; Shugan 1980; Svenson 1979; 1996; Tversky 1972; Wright 1974; overview: Crozier / Ranyard 1997).

alternatives with which the quality of the secondary criteria can be measured. The primary desirability, for valuing actions and the application of secondary desirability criteria, may be a form of desirability prospects (prospective desirability) or it may be a form of total desirability, i.e. a desirability for decisions under certainty (hence a desirability which does without probabilities).

These results are summarized in my second hypothesis:

H2: Primary-and-secondary-desirability hypothesis: Humans use a wide variety of decision strategies implying different criteria of the desirability and optimality of alternatives. These strategies are not inborn but developed in cognitive processes, and their use in different situations is evaluated and chosen according to reflexively optimizing deliberations that imply a primary or fundamental criterion of desirability and optimality.

Of course, this hypothesis does not assume that people are able to formulate their primary or secondary desirability criteria. They must only be able to proceed as if deciding according to such criteria.

The hypothesis just explained again has several advantages. Firstly, it explains why empirical decision researchers could not find very much of what rational decision theory would require for a rational decision - though nevertheless optimizing takes place. In particular most of the decision strategies do not use figures either for probability or for utility. Secondly, even drawing the ultimate conclusion of the application of a decision heuristic may be rather trivial so that it is not internally formulated; in other cases, however, we find thoughts with the content: 'This would be the best.' That means, often there is only an implicit optimality belief but no explicit optimality judgement. This explains the seemingly problem of the optimality belief theory of intention, namely that we rarely find explicit optimality judgements. Finally, the second hypothesis explains the failure of (strongly mathematical) models of decision in the tradition of rational utility theory to find *the* universal desirability criterion used in actual deliberation. There is no such universal desirability criterion but only a wealth of decision strategies implying different desirability criteria.

7. In Search for the Fundamental Desirability Criterion - the Adequacy Condition Hypothesis

My second hypothesis distinguished between secondary and primary or fundamental desirability criteria. A philosophically and psychologically interesting question then is if there is one anthropologically universal primary criterion and, if yes, in what it consists. To my knowledge, until now there is no empirical research trying to answer this specific question. Therefore, I can only speculate about this matter - but speculate on the basis of my experience in rational decision theory and philosophical theory of practical rationality.

In these disciplines criteria for rational desirabilities are discussed and proposed. There is little reason to doubt that many researchers in these disciplines in very fundamental questions decide according to the desirability criteria that they propose and defend. So these criteria for them could be the respective fundamental desirability criteria at the time being. Now there are a lot of controversies between these researchers about the rational desirability criterion. No doubt a big part of these controversies relates to the correct concept of '*intrinsic* desirability' (of consequences or of the action itself) that should be used in the main definition of the fundamental desirability for measuring the value of courses of action. Because the present paper only discusses the latter type of desirability the controversies about 'intrinsic desirability' can be put to one side.[16] But there are controversies about the main definition of 'total desirability' or 'prospective desirability' as well. Many researchers adhere to subjective expected utility theory as the right definition of 'prospective desirability'. Yet even within subjective expected utility theory there are controversies about its correct specification, e.g. which probabilities shall be used: probabilities conditional on the action, probabilities of the circumstances, probabilities of conditionals (with various definitions of these conditionals) or probabilities of causal relations - the different probabilities then lead to different solutions to the Newcomb problem (cf. e.g. Eells 1982; Gibbard / Harper 1978; Jeffrey 1983; Joyce 1999; Lewis 1981; Luce / Krantz 1971). Another problem is how desirabilities of worlds can be computed from intrinsic desirabilities of events, in particular how value independence, absence of intersection and completeness of the events whose desirabilities are to be added can be guaranteed (cf. e.g. Lumer 2000, 292-304; 388-427). And for decisions under uncertainty we still find the rivalry between maximin, maximax, the principle of sufficient reason and other proposals. Other researchers doubt the rationality of some of subjective expected utility theory's prescriptions, in particular that probabilities shall be weighted linearly, and they doubt the axioms from which this follows, too (e.g. Allais 1953; McClennen 1988).

The negative moral of these examples is that there is no anthropologically universal fundamental criterion of desirability. On the other hand researchers discuss the different proposals rationally and they develop new definitions of 'desirability' in a process of technical invention for solving certain problems. Think e.g. of Allais' Paradox - which says that many people in cases of a unique occasion to make a sure and life changing gain will choose this option instead of one with higher expected utility but only probable outcomes. Maximizing expected utility provably is the best strategy in the long run if the law of large numbers holds. Defenders of the so-called paradoxical decision in the Allais situation (i.e. the critics of subjective expected utility theory) underline that in this case the law of large numbers does not hold and that, therefore, a justification for running the risk of gambling away this unique occasion to radically improve one's life for a small increase in expected utility is missing. This means that even the critics of subjective expected utility want to optimize in the long run and that they are looking for technical solutions realizing this aim.

My positive conjecture then is this. Even if there is no anthropologically universal fundamental desirability criterion there may be universally accepted adequacy conditions for deciding between proposals for such desirability criteria. People do not explicitly dispose of such

[16] About the content of intrinsic desires see: Lumer 1997.

adequacy conditions; these conditions only show up implicitly when people decide between different proposals for the fundamental desirability criterion. There are two such conditions of adequacy, one for the fundamental desirability concept to be used in decisions under certainty, the other to be used in other situations. This conjecture is the content of my third and last hypothesis:

H3: Adequacy condition hypothesis: Humans use two fundamental desirability concepts at the same time, a concept of total desirability for decisions under certainty and a concept of prospective desirability for decisions under risk or uncertainty. If subjects choose between such concepts (D_1, D_2, ..., D_n) of fundamental or primary desirability and they believe that D_1 among them comes next to fulfilling the following adequacy conditions they adopt D_1 for their fundamental decisions.

1. Condition for 'total desirability' for decisions under certainty: For all events *x* and *y* with respect to which the subject can decide under certainty holds: *x* (according to the subject's information) is totally better (in the sense of D_i) than *y* exactly if the sum of the intrinsic desirabilities of all intrinsically relevant events accompanying event *x* (according to the subject's information) is higher than the respective sum for *y*.

2. Condition for 'prospective desirability' for decisions without certainty: The desirability criterion D_i in question is materially equivalent (i.e. on the basis of the same information leads to the same preferences) to that desirability criterion D_x (from the set of prospective desirability criteria, which define the prospective desirability of events only on the basis of the subject's empirical information and his intrinsic desirability function) for which holds: if one disregards decision costs, the constant use of D_x as the criterion for decisions without certainty is totally optimum (i.e. optimum according to the fundamental desirability criterion for decisions under certainty).

Both adequacy conditions do not require that a desirability criterion to be chosen has to be identical to one of the two outstanding desirability criteria, namely D_x and the criterion 'sum of the intrinsic desirabilities of all intrinsically relevant events accompanying event *x*', but only that it has to be materially equivalent to one of them. Only this weaker requirement is empirically justified because there are indefinitely many criteria materially equivalent to the outstanding criteria, which with respect to their exactness are equally good. - From the two primary desirability concepts to be defined 'total desirability' is the more fundamental one in that respect that it is already used in the formulation of the adequacy condition for 'prospective desirability'. This at a first glance seems to be rather paradoxical because prospective desirabilities are to be used in situations of lacking certainty only whereas the application of the concept 'total desirability' presupposes complete information. Though the objects to be assessed are different, namely the prospective desirability of singular actions versus the total desirability of the constant use of a certain concept of 'prospective desirability', assessing the total desirability of this constant use seems to imply that one must know the total desirability of all the actions for which the prospective desirability will be assessed, and for which, again, not all relevant consequences are known with certainty. But this is a hasty conclusion. There may e.g. be ways of at least estimating the comparative total desirability of the constant use of a certain concept of 'prospective desirability' that do not require the determination of the total desirability of all the singular actions assessed during that constant use. One such way could be to use the law of large numbers, which says that within a sufficiently large probability sample of objects (e.g. of actions) with probabilistic qualities of a certain type (e.g. having certain consequences) the relative frequency of this quality approaches its probability. Applying this law one may perhaps estimate the total desirability of the constant use of a certain concept of 'prospective desirability' without knowing anything about the outcomes of its singular uses.[17]

The adequacy condition hypothesis can only rarely be used for prognostic purposes. But, supposing it is true, it nonetheless has some important advantages. It can be used for explaining

[17] One objection to the theory just developed is this: Strong psychological principles of decision, as they are assumed by the present theory, allow for predicting decisions and actions. Under certain conditions they even allow for predicting one's own actions. But in such a case decision would be empty. (Shackle 1966, 73 f.; Schick 1979, 237 f.) In addition, Schick tries to show that what seems to be a rare problem is the general case for rational agents: every rational agent must fulfil such conditions which lead him to predict his own action before choosing them and thus emptying his choice (Schick 1979, 239-241).

I do not think that this is really a problem. 1. To begin with the last point first, Schick's argument rests on very strong premises. Most people, including very rational people, e.g. do not know which decision strategy they were following (there is a difference between knowing how and knowing that); and given the multiplicity of decision strategies, in particular stressed by the present theory, they even less know *in advance* which decision strategy they will follow during the impending deliberation. Schick's strongest and I think unreasonably strong premise is that rational agents have to be deductively thorough. This is neither possible nor would it be rational to try to approximate as close as possible to this ideal; the frame problem having arisen in the study of artificial intelligence has elucidated the devasting consequences of this epistemic strategy. 2. The way by which agents arrive at the foreknowledge of their actions, according to Schick's argument, goes via knowing one's options, knowing their consequences, knowing the preference ranking, ..., knowing how one will choose, etc. But working through this process of forming new knowledge up to the point where one comes to believe a certain action to be optimum is *identical* to deliberation, and its last step is identical to deciding. There is no further choice. Hence there is no *fore*knowledge of one's decision either. 3. So Schick's argument has failed. Nonetheless, there may be cases where an agent following a different but still rational way has come to believe that he will do a_1, before having decided to do a_1. But what should be the problem with such cases? 3.1. The determinacy of one's action does not imply that one has no choice in the sense that there are no alternatives open during the decision. '*s* can do a_i', following G. E. Moore's analysis (Moore 1912, ch. 6), should be understood in a conditional way (i.e.: if *s* chooses / decides / intends to do a_i he will do a_i - which ususally holds for many, many options a_i actually not chosen) and not in the sense of a metaphysical possibility. 3.2. Foreknowing one's action likewise does not imply having no alternatives open because again 'disposing of several options' should be understood conditionally and not as attributing a probability above zero to these options. 3.3. But perhaps the problem Shackle and Schick have in mind is still another one. They seem to conceive of a choice problem as answering to the question "Which action shall I take?" (Shackle 1966, 74) or "facing an issue" of "whether to do it [a_1 or a_2 or ...]" (Schick 1979, 238): "Since *s believes* that he will do a_x, he cannot ask himself whether he will." (Schick 1979, 241 [variables changed].) All the formulations just cited are deeply ambiguous, *i.* the first sense always asking for an advisable or preferred option and *ii.* the second sense asking for a prediction. Foreknowing one's action is the answer to the second question. What we are trying to get by deliberation, though, is an answer to the first question. (Only prevision theory sees this differently.) So there is no conflict between foreknowledge and deliberation / decision. Shackle and Schick may have confused the two meanings, thereby implicitely adhering to the prevision theory of intention criticized above. (But even prevision theory can answer to Shackle's and Schick's problem following the lines of criticism 2 above.)

some fundamental decisions people take. And, what is much more important, it can be used in theories of practical rationality for constructing such desirability criteria which are stable with respect to criticisms and getting new information and thus, according to some important philosophical theories of practical rationality, are rational. In other words, if certain desirability concepts can be shown to fulfil these adequacy conditions there is a good chance that reflecting people adopt these desirability concepts as their fundamental desirability concepts and stick to them in the long run even if provided with new information; and this sort of stability would imply their rationality. If this strategy for a theory of practical rationality in the end will be fruitful and lead to strong results has to be found out in independent publications.

So in the end we have found an empirical theory of intention, the optimality-belief theory, which on the one hand reckons with a wide variety of factually adopted conceptions of 'desirability' and thus of 'optimality' but which on the other hand shows a final or aiming point of the rational evolution of such conceptions. The latter, apart from solving some problems in empirical decision theory, could be a rather important result at least for a theory of practical rationality, too, which takes into account what different ways of taking decisions are open to empirical subjects.

Acknowledgement

I would like to thank Robert Audi, Wolfgang Lenzen and Hugh McCann for very helpful discussions of this paper as well as two anonymous referees of "Erkenntnis" for their valuable suggestions.

References

Allais, M. (1953): Le comportement de l'homme rationnel devant le risque. Critique des postulats et axiomes de l'école américaine. In: Econometrica 21. Pp. 503-46.

Audi, Robert (1973): Intending. In: Journal of Philosophy. 70. Pp. 387-402. Reprinted in: Robert Audi: Action, Intention, and Reason. Ithaca; London: Cornell U. P. 1993. Pp. 56-73.

Audi, Robert (1986): Acting for Reasons. In: Alfred R. Mele (ed.): The Philosophy of Action. Oxford [etc.]: Oxford U.P. 1997. Pp. 75-105.

Beach, Lee Roy; Terence R. Mitchell (1978): A Contingency Model for the Selection of Decision Strategies. In: Academy of Management Review 3. Pp. 439-449.

Bechara, Antoine; Hanna Damasio; Daniel Tranel; Antonio R. Damasio (1997): Deciding Advantageously Before Knowing the Advantageous Strategy. In: Science 275. Pp. 1293-1295.

Brand, Myles (1979): The Fundamental Question in Action Theory. In: Noûs 13. Pp. 131-151.

Brand, Myles (1984): Intending and Acting. Toward a Naturalized Action Theory. Cambridge, Mass.; London: MIT Press. vii; 296 pp.

Brandt, Richard B. (1979): A Theory of the Good and the Right. Oxford: Clarendon. xiii; 362 pp.

Bratman, Michael E. (1987): Intention, Plans, and Practical Reason. Cambridge, Mass.; London: Harvard Univ. Pr. xi; 200 pp.

Camerer, Colin F. (1992): Recent Tests of Generalizations of Expected Utility Theories. In: Ward Edwards (ed.): Utility theories. Measurement and Applications. Boston; Dordrecht; London: Kluwer. Pp. 207-251.

Camerer, Colin [F.] (1995): Individual Decision Making. In: John H. Kagel; Alvin E. Roth (eds.): The Handbook of Experimental Economics. Princeton, NJ: Princeton University Press. Pp. 587-703.

Churchland, Paul M. (1970): The Logical Character of Action-Explanations. In: Philosophical Review 79. Pp. 214-236.

Crozier, Ray [W.]; Rob Ranyard (1997): Cognitive process models and explanations of decision making. In: Rob Ranyard; W. Ray Crozier; Ola Svenson (eds.): Decision Making. Cognitive Models and Explanations. Oxford: Routledge. Pp. 3-20.

Currim, Imran S.; Rakesh K. Sarin (1989): Prospect versus Utility. In: Management Science 35. Pp. 22-41.

Davidson, Donald (1963): Actions, Reasons, and Causes. In: Donald Davidson: Essays on Actions and Events. Oxford: Oxford U.P. 1980. Pp. 3-19.

Davidson, Donald (1978): Intending. In: Donald Davidson: Essays on Actions and Events. Oxford: Oxford U.P. 1980. Pp. 83-102.

Donagan, Alan: Choice (1987). The Essential Element in Human Action. London; New York: Routledge & Kegan Paul. x; 197 pp.

Eells, Ellery (1982): Rational Decision and Causality. Cambridge: Cambridge U.P. x; 234 Pp.

Gibbard, Allan; William L. Harper (1978): Counterfactuals and Two Kinds of Expected Utility. In: C. A. Hooker; J. J. Leach; E. F. McClennen (eds.): Foundations and Applications of Decision Theory. Vol. I. Dordrecht: Reidel. Pp. 125-162. - Reprinted in: Peter Gärdenfors; Nils-Eric Sahlin (eds.): Decision, Probability and Utility. Selected Readings. Cambridge: Cambridge U.P. 1988. Pp. 341-376.

Gigerenzer, Gerd; Peter M. Todd; the ABC Research Group (1999): Simple Heuristics That Make Us Smart. Oxford [etc.]: Oxford U.P. xv; 416 pp.

Ginet, Carl (1990): On action. Cambridge [etc.]: Cambridge U.P. ix; 159 pp.

Goldman, Alvin I. (1970): A Theory of Human Action. Englewood Cliffs, New Jersey: Prentice-Hall. x; 230 pp.

Gollwitzer, Peter M. (1991): Abwägen und Planen. Bewußtseinslagen in verschiedenen Handlungsphasen. Göttingen; Toronto; Zürich: Hogrefe. XV; 254 pp.

Gollwitzer, Peter M. (1996): Das Rubikonmodell der Handlungsphasen. In: Julius Kuhl; Heinz Heckhausen (eds.): Motivation, Volition und Handlung. = Enzyklopädie der Psychologie. Themenbereich C, Serie IV, Bd. 4. Göttingen [etc.]: Hogrefe. Pp. 531-582.

Gollwitzer, Peter M. (1999): Implementation Intentions. Strong Effects of Simple Plans. In: American Psychologist 54. Pp. 493-503.

Habermas, Jürgen (1975): Handlungen, Operationen, körperliche Bewegungen. In: Jürgen, Habermas: Vorstudien und Ergänzungen zur Theorie des kommunikativen Handelns. Frankfurt, Main: Suhrkamp. 1984. Pp. 273-306.

Habermas, Jürgen (1981): Theorie des kommunikativen Handelns. Bd. 1: Handlungsrationalität und gesellschaftliche Rationalisierung. Frankfurt, Main: Suhrkamp. 534 pp.

Harless, David W.; Colin F. Camerer (1994): The predictive utility of generalized expected utility theories. In: Econometrica 62. Pp. 1251-1289.

Harman, Gilbert (1976): Practical Reasoning. In: Review of Metaphysics 29. Pp. 431-463.

Heckhausen, Heinz (1987): Intentionsgeleitetes Handeln und seine Fehler. In: Heinz Heckhausen; Peter M. Gollwitzer; Franz E. Weinert (eds.): Jenseits des Rubikon. Der Wille in den Humanwissenschaften. Berlin [etc.]: Springer. Pp. 143-175.

Heckhausen, Heinz (^{2}1989): Motivation und Handeln. Zweite, völlig überarbeitete und ergänzte Auflage. Berlin [etc.]: Springer. xviii; 557 pp.

Heckhausen, Heinz; Peter M. Gollwitzer; Franz E. Weinert (eds.) (1987): Jenseits des Rubikon. Der Wille in den Humanwissenschaften. Berlin [etc.]: Springer. xiv; 420 pp.

Jeffrey, Richard C. (^{2}1983): The Logic of Decision. (1965.) 2nd Edition. Chicago; London: University of Chicago Press. xiv; 231 pp.

Johnson, E. (1979): Deciding how to decide: The effort of making a decision. Unpublished manuscript: University of Chicago.

Joyce, James (1999): The Foundations of Causal Decision Theory. Cambridge: Cambridge U.P. 280 pp.

Kahneman, Daniel; Amos Tversky (1979): Prospect theory. An analysis of decision under risk. In: Econometrica 47. Pp. 263-291.

Kambartel, Friedrich (1974): Moralisches Argumentieren. Methodische Analysen zur Ethik. In: Friedrich Kambartel (ed.): Praktische Philosophie und konstruktive Wissenschaftstheorie. Frankfurt: Suhrkamp. Pp 54-72.

Lennon, Kathleen (1990): Explaining Human Action. London: Duckworth. 176 pp.

Lewis, David (1981): Causal decision theory. In: David Lewis: Philosophical Papers. Vol. II. New York; Oxford: Oxford U.P. 1986. Pp. 305-337 (337-339).

Luce, R[obert] D[uncan]; D. H: Krantz (1971): Conditional Expected Utility. In: Econometrica 39. Pp. 253-271.

Lumer, Christoph (1990): Praktische Argumentationstheorie. Theoretische Grundlagen, praktische Begründung und Regeln wichtiger Argumentationsarten. Braunschweig: Vieweg. XI; 474 pp.

Lumer, Christoph (1997): The Content of Originally Intrinsic Desires and of Intrinsic Motivation. In: Acta analytica - philosophy and psychology 18. Pp. 107-121.

Lumer, Christoph (forthcoming): Kognitive Handlungstheorie. Empirische Handlungsgesetze, Freiheit und die Grundlagen praktischer Rationalität.

Maslow, Abraham H[arold] (1954): Motivation and Personality. 2nd ed. New York: Harper and Row ^{2}1970. xxx; 369 pp.

McCann, Hugh (1998): Practical Rationality and Weakness of the Will. In: Hugh McCann: The Works of Agency. On Human Action, Will, and Freedom. Ithaca; London: Cornell U.P. Pp. 213-233.

McClennen, Edward F. (1983): Sure-thing doubts. In: Peter Gärdenfors; Nils-Eric Sahlin (eds.): Decision, Probability and Utility. Selected Readings. Cambridge: Cambridge U.P. 1988. Pp. 166-182.

Mead, George Herbert (1934): Mind, Self and Society. From the Standpoint of a Social Behaviorist. Edited and with an introduction by Charles W. Morris. Chicago; London: University of Chicago Press ^{18}1972. xxxviii; 401 pp.

Mele, Alfred R. (1992): Springs of Action. Understanding Intentional Behavior. New York; Oxford: Oxford U.P. xi; 272 pp.

Moore, G[eorge] E[dward] (1912): Ethics. London [etc.]: Oxford U.P. ^{2}1966. 137 pp.

Norman, Donald A. (1981): Categorization of Action Slips. In: Psychological Review 88. Pp. 1-15.

Payne, John W.; James R. Bettman; Eric J. Johnson (1988): Adaptive Strategy Selection in Decision Making. In: Journal of Experimental Psychology: Learning, Memory and Cognition 14. Pp. 534-552.

Payne, John W.; James R. Bettman; Eric J. Johnson (1993): The adaptive decision maker. Cambridge: Cambridge U.P. xiii; 330 pp.

Reber, Arthur S. (1989): Implicit Learning and Tacit Knowledge. In: Journal of Experimental Psychology, General. 118. Pp. 219-235.

Rescher, Nicholas (1959/60): Choice without Preference. A Study of the History and of the Logic of the Problem of "Buridan's Ass". In: Kant-Studien 51. Pp. 142-175.

Rubinstein, Ariel (1988): Similarity and Decision-making under Risk. Is There a Utility Theory Resolution to the Allais Paradox? In: Journal of Economic Theory 46. Pp. 145-153.

Schick, Frederic (1979): Self-knowledge, Uncertainty and Choice. In: The British Journal for the Philosophy of Science 30. Pp. 235-252.

Sen, Amartya (1997): Maximization and the Act of Choice. In: Id.: Rationality and Freedom. Cambridge, Mass.: Harvard U.P. 2002. Pp. 158-205.

Shackle, G. L. S. (1966): The Nature of Economic Thought. Selcted Papers 1955-1964. Cambridge: Cambridge U.P. xiv; 322 pp.

Shugan, S. M. (1980): The cost of thinking. In: Journal of Consumer Research 7. Pp. 99-111.

Simon, Herbert A. (1956): Rational choice and the structure of the environment. In: Psychological Review 63. Pp. 129-138.

Slote, Michael (1989): Beyond Optimizing. A Study of Rational Choice. Cambridge, Mass.; London: Harvard Univ. Pr. xi; 192 pp.

Slovic, Paul; Sarah Lichtenstein; Baruch Fischhoff (1988): Decision making. In: Richard C. Atkinson; Richard J. Herrnstein; Gardner Lindzey; R. Duncan Luce (eds.): Steven's handbook of experimental psychology. Vol. 2: Learning and cognition. New York [etc.]: Wiley. Pp. 673-738.

Svenson, Ola (1979): Process Descriptions of Decision Making. In: Organizational Behavior and Human Performance 23. Pp. 86-112.

Svenson, Ola (1996): Decision Making and the Search for Fundamental Psychological Regularities. What Can Be Learned from a Process Perspective? In: Organizational Behavior and Human Decision Processes 65. Pp. 252-267.

Tversky, Amos (1972): Elimination by aspects. A theory of choice. In: Psychological Review 79. Pp. 281-299.

Tversky, Amos; Daniel Kahneman (1992): Advances in Prospect Theory. Cumulative Representation of Uncertainty. In: Journal of Risk and Uncertainty 5. Pp. 297-323.

Ullmann-Margalit, Edna; Sidney Morgenbesser (1977): Picking and Choosing. In: Social Research 44. Pp. 757-785.

Velleman, J[ames] David (1989a): Practical Reflection. Princeton, N.J.: Princeton, U.P. x; 332 pp.

Velleman, J[ames] David (1989b): Epistemic Freedom. In: Pacific Philosophical Quarterly 70. Pp. 73-97.

Wright, Georg Henrik von (1963): Practical Inference. In: The Philosophical Review 72. Pp. 159-179.

Wright, Georg Henrik von (1971): Explanation and Understanding. London: Routledge & Kegan Paul. xvii; 230 pp.

Wright, Georg Henrik von (1972): On So-Called Practical Inference. In: Acta Sociologica 15. Pp. 39-53.

Wright, Peter [L.] (1974): The Harassed Decision Maker. Time Pressures, Distractions, and the Use of Evidence. In: Journal of Applied Psychology 59. Pp. 555-561.

Università degli Studi di Siena
Dipartimento di Filosofia
Via Roma, 47
I-53100 Siena
Italy

E-mail: lumer@unisi.it