An Empirical Theory of Practical Reasons and Its Use for Practical Philosophy

Christoph Lumer

(University of Siena)

Abstract: In the first part (2-5) an empirical theory of practical reasons is sketched and defended. It consists of: hypotheses about what intentions are, namely optimality beliefs, (2), hypotheses about how intentions are formed on the basis of probabilistic beliefs and intrinsic desires (3), a pluralist theory about intrinsic desires (4) and a theory about motives for moral action (5). In the second part (6-8) it is argued that normative practical philosophy must rely on empirical theories of practical reasons (6). Then it is shown that theories of prudential practical rationality and of morals that are not based on such empirical theories run into problems. And it is sketched how the empirical theory outlined in the first part can be used as a basis for a reconstructive approach to practical rationality and of an internalist approach to morals (7-8).

1. The aims of this paper and the concept of 'practical reason'

Theoretical reasons are reasons for believing, practical reasons are reasons for acting. Two main types of practical reasons have to be distinguished. 1. *Normative reasons* - which have to be differentiated again into 1.1. rational or prudential and 1.2. moral reasons - are good reasons, which serve to choose between actions (and to justify them later on) in a rational, prudential or moral way. Ethics and the theory of practical rationality imply theories about normative reasons. 2. *Explanatory reasons* are reasons for which someone acted; they explain why the person acted as he did. They are dealt with in philosophy of action.[1]

In practical philosophy nowadays there is much talk of "practical reasons", their ontology, their content, their rationality, their intelligibility, if and how they motivate etc. One problem of these debates is that the normative theories often make or presuppose strong assumptions about the empirical features of our explanatory practical reasons but that these assumptions are almost never proved and developed into a coherent theory, so that the normative theories remain unfounded in this respect.[2]

---

[1] This is a usual distinction [see e.g. Nagel 1970, 4; 18; Smith 1987, 38 f.; Dancy 2000, 1-4], though the explanatory reasons sometimes are called "motivating reasons". Holding that normative reasons should be motivating, too, I prefer to speak of "explanatory" versus "normative reasons".

[2] Examples are Dancy's attempts to prove cognitivism and to prove that all explanatory reasons are objective facts [Dancy 2000, 85; 98-116; 126-133]. These endeavours are criticized in: Lumer 2004a, sect. 4. Another example is Searle's "proof" that there are desire-independent motivating reasons. At the crucial point he simply assumes that recognizing one's social obligation is already motivating [Searle 2001, 170 f.; 181] - without providing even a trace of an argument or explanation for this assumption.

This paper has two aims and parts. It first will sketch a *theory* of explanatory reasons, i.e. an empirical decision theory about how people deliberate, decide and form intentions, culminating in several empirical hypotheses (sects. 2-5). (This part to a great extent summarizes pieces of my research. References will be provided in the single sections.) Subsequently it will be explained why and how normative theories of practical reasons depend on this type of empirical information (sect. 6); and some exemplary consequences of the empirical theory for normative theories of practical reasons will be drawn, namely for the theory of practical rationality and for ethics (sects. 7-8).

The mentioned boom of talking about "practical reasons" has a further problematic aspect. This talk is often very unclear and vague; sometimes it is even intended to be so. But there is also a cause to the point of this vagueness: 'Explanatory reason' is only a comprehensive notion, a variable for all kinds of considerations that have influenced the final intention. (In ordinary language we can call e.g. any of the following instances that contributed to the decision an "(explanatory) reason" for the action: desires, criteria underlying such desires, (probabilistic) beliefs about means-ends-relations, about other implications of the action, about the set of alternatives, comparative valuations of actions, beliefs about events that have caused certain desires, about circumstances that constitute particular occasions for actions, beliefs about facts that may justify the beliefs listed so far, beliefs about authority arguments for the beliefs listed so far. And if the speaker thinks the agent's beliefs to be true he may also call the respective states of affairs "reasons". [Lumer 2004a, 709 f.]) So '*explanatory reason* for an action *a*' may best be defined as: (self-contained) part of the relevant considerations and mental attitudes in the deliberation leading to the intention to *a*, where this intention has caused *a* in the right way [cf. Lumer 2004a, 710].[3] The empirical theory to be sketched in the following (sects. 2-5) is also an effort to explain for what kinds of mental attitudes the variable 'explanatory reason' empirically stands.

---

[3]   Some words about the ontology of reasons: Whereas normative reasons have an objective form - 'that the shares of IBM will increase in price tomorrow is a normative reason to buy them today' -, explanatory reasons are presented in a subjective as well as in an objective form: 'She bought IBM because she thought these shares would increase the following day' versus 'She buys IBM because these shares will rise tomorrow.' The classical explanation of this discrepancy is: the correct and primary form of explanatory reasons is subjective; but we can and should use the objective form when we have a justified belief that this reason is true; the speaker thereby gives additional information. Explanatory reasons are subjective because they report the considerations and propositional attitudes from which the agent decided and acted and by which the action shall be explained - though the doxastic explanatory reasons aim at truth, i.e. they try to capture objective facts, which are the normative reasons. Explanatory reasons cannot be primarily objective just because the states of affairs in question in part are causal relations (and not events), happen in the future or are not facts and therefore cannot explain the action. (This ontology is neutral with respect to the question of what is foundationally primary, normative reasons, as externalism and in particular value objectivism holds, or explanatory reasons, as internalism maintains.)

2. The decisive last mental step before action: intentions and what they are [4]

The desire-belief theory of action holds that action arises from pairs of desires for some end and beliefs that a certain action will realize that end. Qualifications of this idea have been, for example, that actions arise from the strongest desire or the desire-belief pair with the highest product of desirability and probability etc. [cf. e.g. Davidson 1963, in particular 3-12; 1978, 87; Goldman 1970, 49-57; 72]. The desire-belief theory is problematic because it tries to do without intentions, thus underrating their role which cannot be fulfilled by sets of desire-belief pairs [Bratman 1987]: 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. Therefore, the immediate and decisive mental cause of actions must be intentions.

There is a host of theories about what propositional attitudes intentions are. The most important ones are: *sui-generis theory*, which says 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]; *prevision theory*, which maintains that intentions are - self-fulfilling - previsions of one's own actions [Harman 1976; Velleman 1989a, 109 and 109-142; 1989b]; *optimality belief theory*, according to which intentions are beliefs that among the considered set of open alternatives, one specific course of action is optimum [Lumer 2005];[5] *models of practical inference* [Aristotle, NE 1111b-1113b; 1139a; 1147a; von Wright 1963; 1971, 96-107; 1972]; *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]; *models of needs*, presuming that we act out of our strongest or most developed need and a belief that a certain action will fulfil that need [Maslow 1954, chs. 4-7].

The majority of these theories are *doxastic theories of intention*, holding that intentions are certain types of beliefs that the intended act token (I will perform an action of type *A* at time *t*) has a certain property *F*. They differ in their respective assumptions about that *F*. *Prevision theory* holds that *F* is empty (or equal to 'is true' - the intention then is the belief 'I will do *A* at *t*' or 'It is true that I will do *A* at *t*'); *psychological normativism* holds that *F* is equal to 'is socially required or prescribed', etc. Further doxastic theories about intention are the optimality-belief theory, models of practical inference and perhaps (depending on the exact interpretation) models of need.

---

[4]   This and the following section are based on: Lumer 2005.

[5]   Somewhat more precisely, the optimality-belief theory holds: Forming an intention consists in an optimality judgement that a certain action is optimum among the considered set of open alternatives. And the intention is the resulting optimality belief. But they are intentions only up to the 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 person 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. [Lumer 2005, sect. 4.]

Intentions have many features; in the following only important empirical features apt to falsify some theory of intention are considered. Doxastic theories of intention as such have some very important advantages in the respect that they can easily explain the following features.

I1. Present and distal intentions: There are present and distal intentions. Doxastic theories explain this difference easily with the action description's differing time indices.

I2. Various logical forms of intentions: Intentions come in various logical forms; there are e.g. general, conditional or disjunctive intentions. Doxastic theories explain this by assuming that such intentions are of the form 'all *a* (of that and that type) are *F*', 'if *c* then *a* is *F*', '*a* is *F* or *b* is *F*' and that we may arrive at elementary intentions via logical inferences.

I3. Transition deliberation to intention: Usually there is an easy transition from the deliberation, which may precede the intention, to intending. Doxastic theories easily explain this by assuming that the deliberation in such cases consists in a cognitive process during which a true proposition of the intention's form is sought and that the intention formation consists in, finally, after having found sufficient reasons, to come to believe in the respective proposition.

I4. Settling effect of intentions: Intentions, to a certain degree, settle what will be done and create a commitment to the execution,[6] which, though, may eventually be withdrawn later. Doxastic theories explain this so: forming an intention, at least temporarily, stops the consideration of alternatives because after pondering various pros and cons the person finally comes to believe that *a* is *F*. And withdrawal of the commitment is due to finding new reasons that lead to a change in one's opinion.

I5. Small and big actions: Human intentions aim at a very wide spectrum of intended actions, e.g. from scratching one's knee up to doing what one can to diminish hunger in the world. And often small actions are intended as parts of bigger actions. Doxastic theories obviously allow "small" as well as "big" actions to be the object to which the predicate '*F*' may be applied. And they explain the transition from intending a big action to intending its parts this way. There are several important candidates for '*F*' for which holds: if *F* applies to a big action then *F* applies to all its (relevant) parts, too, so that the coming into existence of the small intentions can follow the paths of logic.

However, apart from optimality belief theory, none of the other doxastic theories (prevision theory, psychological normativism, models of practical inference and models of needs) can explain a further highly important feature of intentions, namely this:

I6. The weighing content of deliberation: The most important content of deliberation, which might precede the formation of intention, is to ponder the possible advantages and disadvantages of the considered 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

---

6   Bratman has emphasized this settling effect as a prerequisite for intrapersonal and interpersonal coordination [Bratman 1987, 4; 16 f.].

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. In none of these cases is pondering pros and cons of importance.

In addition, psychological normativism, models of practical inference and models of needs (but not prevision theory and again not the optimality belief theory) fail to explain the uniqueness of intended actions.

I7. Definiteness of intentions: In executive intentions we choose exactly *one* action as the one to be executed (or a well circumscribed set of actions about which the agent is undecided and between which he lets the executive system decide as a kind of random device). This definiteness is necessary for the executive system to "know" what to do.

All the restricting mechanisms appealed to by psychological normativism, models of practical inference and models of needs, i.e. social norms, aims and needs leave much room for various alternatives. On the other hand optimality belief theory can easily explain the latter two features of intention: I6. Pondering pros and cons of an isolated action means considering what makes this action better or worse, respectively. Considering the pros and cons of one action as compared to those of another action means trying to find out, which is the better one. I7. Uniqueness of optimality beliefs results from restricting the set of candidates of possible actions via the relation 'is better than'. ('a_1 is optimum among the set of options $\{a_1, a_2, ..., a_n\}$' means: a_1 is better than all the other options, i.e. $a_2, a_3, ..., a_n$.) In case of options with utilities close to each other definiteness may be reached by fine-tuning the options' valuation. The limiting cases then are Buridan examples where we end up with two (or more) options held to be exactly equally desirable. The relation 'is better than' does not help any longer in this situation. But instead of falsifying optimality belief theory, such cases even confirm it because optimality belief theory explains the difficulties we actually have in real life with such situations; we still search for the better action. (A way out for cases where we ultimately do not find any value difference is to, finally, form a disjunctive intention and thus let the executive system decide between the weakly best options. From the intention system's viewpoint this comes up to a random decision.)

The currently most famous *non*-doxastic theory of intention is the *sui-generis theory*, which says that intentions are irreducible propositional attitudes. Apart from the fact that introducing a new propositional attitude seems to be a bit *ad hoc*, such theories have some difficulties in explaining the above mentioned features I1, I2, I3, I5 and I6 of intentions. These difficulties are not insurmountable in all cases but every condition requires an extra explanation and an additional, complicating part of the theory - only Mele has offered such an explanation and only for features I3 and I6 [Mele 1992, 228-234]. The sui-generis theory e.g. must hold that conditional intentions (cf. I2) are attitudes of type   with the propositional content 'if *p* then I will do *A* at *t*'; and it must explain why, after coming to believe that *p*, the person develops an attitude of type   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 features I1 (present and distal intentions) and I5 (small and big actions). But above all, the sui-generis model of intention does extremely badly with feature I3 (transition from deliberation to intention) and, consequently, with

feature I6 (the weighing content of intentions). According to the sui-generis model, the intention is quite isolated from the deliberation. Mele has tried to resolve this problem by complementing the sui-generis model with a theory about the acquisition of intentions [Mele 1992, 228-241]. According to this theory, the most important cause of intentions are optimality judgements about actions, which by default, i.e. in the absence of preemption, lead to the respective intention [ibid. 228-234]. However if this were so then the optimality belief itself could function as an intention, as is assumed by the optimality belief theory, and the *sui-generis* intention would be superfluous.

Summarizing, we may conclude that if there are no further very important features of intentions or further promising explanations of intentions not yet considered the optimality-belief theory is the best explanatory theory of intention.

H1: Intentions as optimality beliefs: Intentions are optimality beliefs that a certain action is optimum (in a sense to be specified) among the considered set of options.

However there are some well-known objections to the optimality-belief theory of intention. Here I can deal only with the most important of them: the *phenomenological objection*.[7] It says, considering the content of our deliberation and decisions introspectively we only rarely find more or less explicit thoughts of the content 'This (action, option) would be best.' In particular tiny children, when they are just barely able to intend, probably do not dispose of the concepts figuring essentially in optimality propositions [Robert Audi, personal communication]. This problem of the interpretation of optimality propositions will be dealt with in the next section.

3. Forming intentions: aggregating intrinsic valuations to optimality judgements

The optimality-belief theory introduced so far 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. The second part of the empirical theory of practical reasons to be outlined here deals with how the final optimality judgement is aggregated from more basic valuations, thereby taking into account various consequences of the actions considered and the different probabilities of these consequences.

But first some terminology has to be explained. 1. I speak of '*desirabilities*' instead of 'utilities' because the latter notion is bound to a very specific theory of empirical desirabilities, namely subjective expected utility theory. 2. *Empirical desirabilities* are the desirabilities people have in mind during deliberation. Ideally such desirabilities are quantitative. But usually people have only vague ideas about the respective quantity; they may perhaps use comparisons to other quantities but not numbers. These ideas about quantities, nonetheless, go beyond mere comparative judgements, i.e. subjects can say not only that a_1 is better than a_2 but also how much better it is. 3. Apart from empirical desirabilities there are *rational desirabilities* of objects, which follow from empirical qualities of that object and the definition of 'desirability' given in rational or prudential

---

7   Some further objections are discussed in: Lumer 2005, sect. 4.

desirability theories. The present part of this paper deals with empirical desirabilities only. 4. The *total desirability* of an object *a* is the desirability of *a* under the condition that all its relevant consequences are known with certainty. The total desirability is equal to *a*'s intrinsic desirability plus the intrinsic desirabilities of *a*'s consequences. 5. The *prospect desirability* of an object *a*, on the other hand, can also be determined when not all relevant consequences of *a* are known with certainty. Apart from the intrinsic desirability of *a* itself *a*'s prospect desirability includes weighted desirabilities of *a*'s (more or less probable) consequences. There are various ways of weighting possible consequences; the most prominent is to weight them according to their expected desirability. So expected desirability and total desirability are special cases of prospect desirability.

The idea behind prospect desirabilities is to summarize everything that has some more basic desirability and probably accompanies *a* into *one* judgement about *a*. The rationale behind this is that in this way some sort of index is generated saying a how good world will accompany *a*. The concept of 'total desirability' does the same for certain prospects.

According to the meaning of "Action a_1 is optimum (i.e. has the highest prospect desirability) among the options $\{a_1, a_2, ..., a_n\}$ taken into consideration", such judgements should be verified by, first, assessing the single options, attributing to each of them a prospect desirability, and, second, by comparing the prospect desirabilities of all options to find out which is highest. In practice however we find little of this ideal procedure. There is no straightforward chronological order of the single steps, numbers are almost never used, and the procedure is greatly abridged.

Empirical decision theory, which is mainly advanced by psychologists and economists, deals with the problem of aggregating more basic desirabilities to decisions. There are three main approaches. They all say that decision consists in considering the various pros and cons of options, i.e. their expected positive or negative consequences, and in aggregating the respective probability assumptions and desirabilities of the consequences. But these theories differ in their hypotheses about how this is done exactly. The first one is *subjective expected utility theory*, which says that we arrive at decisions *as if* they had been taken according to the advices of rational decision theory, i.e. by calculating the expected utility of the options taken into consideration and choosing one of the options with the highest expected utility. Because rational decision theory is a *normative* theory it would be very astonishing if all persons complied with that theory. Actually psychologists have falsified subjective expected utility theory in many respects, especially since the late 1960s [e.g. Kahneman & Tversky 1979; overview: Camerer 1995, 622-626; 644-649; 652-670]. It has been shown e.g. that the great majority of subjects does not weight risky prospects linearly, i.e. multiplying the desirability of the respective outcome with its probability. Highly probable but not certain prospects are undervalued, whereas prospects with little probabilities are overvalued.

The second type of empirical decision theories are axiomatic theories of empirical desirabilities a bit in the fashion of subjective expected utility theory but generalizing this approach in allowing for interpersonally differing ways of non-linear weighting of not certain (i.e. risky or uncertain) prospects. This means these theories provide only a general formula of risk weighting, and the exact weighting function for risky prospects, in general a non-linear function, has to be

established for each person separately. The most prominent model of this kind is Kahneman's and Tversky's prospect theory [Kahneman & Tversky 1979; Tversky & Kahneman 1992] but there are many others [overview: Camerer 1995, 626-644].

However all the axiomatic or general theories of empirical desirabilities developed so far have two big problems. First, even though some of these theories are elegant and many of them mathematically rather ambitious several studies have found that none of them is prognostically satisfying [cf. e.g. Harless & Camerer 1994; Currim & Sarin 1989]. In normal cases they are not prognostically better than subjective expected utility theory. This holds even for prospect theory. Second, those axiomatic theories as well as subjective expected utility theory are *hydraulic* (and not cognitive) in the sense of not pretending to model what people actually think during deliberation but only to model their final decision based on their assumptions about desirabilities and probabilities, which in some way and beyond conscious awareness then leads to the predicted decision. Most people would not even understand the formulas used for description. 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.

The third approach in 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.[8] Up 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 general idea crystallizes from these findings. The general aim of deliberation is to find the best option - but with some qualifications. First, the aim of finding the best option does not require calculating in the ideal form described above, i.e. calculating the prospect desirability of each option and then comparing these values. There are many possible abbreviations, which do not even change the result. If e.g. option a_{n+1} is equal to a_n in all but one respects and worse in that particular respect, then a_{n+1} is worse than a_n; and if a_{n+1} is found to be worse than a_n it cannot be optimum and its prospect desirability need not to be calculated. Second, estimating prospect desirabilities may be more or less precise, depending on the inclusion of more or less important aspects of the object. A very rough deliberation in the end may regard only one important aspect of the option but still provide an approximation of the option's prospect desirability. Third, options are more or less important. The cost of possible errors due to less precise assessments (i.e. the costs of not having chosen the really best option but only the, let us say, third best option) diminish with the importance of the options. Fourth, deliberation is costly, and its costs have to be added to those of the action itself. On the other hand, deliberation may have (perhaps big) advantages, namely to lead to finding a better option. So we have a second problem of optimization here, that of optimizing deliberation. Although this kind of optimizing cannot be very precise (the attempt to do so would lead to more and more costly higher order optimizations

---

8   Some important contributions of this type are e.g.: Gigerenzer et al. 1999; Payne et al. 1993; Shugan 1980; Svenson 1996. Overview: Crozier & Ranyard 1997.

up to infinite levels) it is possible to use rules of thumb that say how much one should deliberate in case of a certain sort of decision problem.

Now at least adults are implicitly aware of all this and somehow take it into account in that they "decide" how to decide, i.e. they change their decision strategies intrapersonally with the type of decision problem, in the end aiming at optimizing deliberation, thereby optimizing the executive action as well [cf. e.g. Payne et al. 1993; Shugan 1980; Svenson 1996]. The observable (via process tracing methodologies) result of this underlying tendency is the huge and at first sight confusing multitude of decision strategies used. The simplest strategy is to choose an option immediately when it comes into mind if it has at least one positive aspect and no (grossly) negative aspect of it comes to mind. This decision strategy is often used in real life, it is open even to tiny children, and in philosophy it has been formalized as practical syllogism. This implicit qualification of the optimality-belief theory should suffice for rejecting the phenomenological objection that in deciding we usually do not think in terms of 'optimality' (cf. above, sect. 2).

Now if behind the great variety of decision strategies used there is a tendency of optimizing deliberation then the following must hold. During deliberation about executive actions people use many decision strategies implying *many secondary* criteria of the desirability and optimality of the options. These secondary criteria are not anthropologically fixed; they are not even intrapersonally fixed but have been cognitively developed. However, on a deeper level there seems to be *one primary*, i.e. fundamental and most exact, criterion of the desirability of options with which the quality of the secondary criteria can be measured. The primary desirability may be a form of prospect desirability or it may be a form of total desirability. Of course, people in most cases will not be able to formulate this primary criterion; they need to know it only implicitly.

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

If secondary desirability criteria are not anthropologically fixed, at least primary criteria may be so. Until today there has been no empirical research about this question. But the fact that even decision theorists, who should be able to formulate their most fundamental desirability criterion, strongly disagree about many aspects of such a criterion makes quite unlikely that there is such an anthropologically fixed primary desirability criterion. On the other hand the discussion among them is not simply intuitionistic and does not only consist of diverging statements but includes arguments. And it leads to new definitions of 'desirability' in a process of technical invention for solving certain problems. A weaker but still anthropologically universal hypothesis explaining these developments then is this. There is no universally accepted primary desirability criterion; but there are universally accepted *adequacy conditions* for deciding between proposals for fundamental desirability criteria. People do not explicitly dispose of such adequacy conditions; these conditions only show up implicitly when people decide between different proposals for the

fundamental desirability criterion. Observing the proposals made during the discussion, one may formulate two adequacy conditions underlying them:

H3: *Adequacy conditions for fundamental desirability concepts:* Humans use two fundamental or primary desirability concepts at the same time, a concept of total desirability for decisions under certainty and a concept of prospect desirability for decisions under risk or uncertainty. If humans choose between concepts $(D_1, D_2, ..., D_n)$ of fundamental desirability and they believe that among them D_1 comes closest to fulfilling the following adequacy conditions whereas the other do not 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 person can decide under certainty holds: *x* (according to the person's information) is totally better (in the sense of D_i) than *y* iff the sum of the intrinsic desirabilities of all intrinsically relevant events accompanying event *x* is higher than the respective sum for *y*.
2. Condition for 'prospect desirability' for decisions without certainty: The desirability criterion D_i in question is materially equivalent (i.e. on the basis of the same information it leads to the same preferences) to that desirability criterion D_x (from the set of prospect desirability criteria, which define the prospect desirability of events only on the basis of the person'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). [Cf. Lumer 2005, 252.]

This adequacy condition hypothesis can only rarely be used for prognostic purposes. But it nonetheless can be used for explaining 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 in view of criticism and of getting new information. Such stability then may be considered as a guarantee of their rationality.

4. Intrinsic valuations

4.1 Conceptual distinctions - what are originally intrinsic desires?

The second part of the empirical theory of decision sketched here (sect. 3) dealt with the aggregation of more basic desirabilities of actions themselves and their consequences as well as information about the probabilities of such consequences to optimality judgements about actions in terms of 'total' or 'prospect desirabilities'. The more basic desirabilities of consequences may again be prospect desirabilities or total desirabilities. But in order not to lead to an infinite regress there must be some more basic desirabilities that are not prospect or total desirabilities [cf. Aristotle, NE 1094a18-22]. These most basic desirabilities are *intrinsic desirabilities*, i.e. desirabilities that we do not attribute to the object because of its consequences but because we are interested in some feature

of the object itself.[9] The more basic desirabilities for actions themselves are always immediately intrinsic - otherwise the more basic desirability would have to include again some of the action's consequences, which already are considered separately. Intrinsic desirabilities express what we are interested in properly. Total and prospect desirabilities only summarize such intrinsic desirabilities for facilitating decisions.

Even some prospect desirabilities are more basic than others in that the first are the desirabilities of the latter's consequences; so, for example the desirabilities of having food will usually be more basic than the desirability of having the money to buy it. However even the desirability of having food is still a prospect desirability, and having food itself is probably intrinsically neutral. Assessing intrinsic desirabilities sometimes seems to be rather trivial; at least adults aggregate them (nearly) automatically to some rather basic prospect desirabilities. Therefore, most desirabilities, even the most basic ones, we are *explicitly* considering during deliberation are already *prospect* desirabilities. So most people do not even differentiate explicitly between intrinsic and prospect desirabilities. Nonetheless, we can find out people's intrinsic desirabilities by asking them whether they do not think that they desire some object because of its consequences. And we may invite them to reflect if their valuation perhaps might change if they imagined for a moment that the object in question would not have the assumed consequences. Another way of finding out people's intrinsic desirabilities is deep psychology, which historically reconstructs how people come to value certain objects in a specific way. Because these two methods may lead to different results we should distinguish between *originally intrinsic desires / desirabilities*, which should be the aim of deep psychological research, and (at time *t*) *conscious intrinsic desires / desirabilities*, which are those intrinsic desires that could be discovered by means of the first method. We may find conscious intrinsic desires that originally were prospect desires only. And in case of repression we may even find originally intrinsic desires that presently have no conscious counterpart at all.

The difference between intrinsic desirabilities and prospect desirabilities is not only mostly neglected in usual deliberation, even standard utility theory and rational decision theory do not make this distinction as they only look for coherence of unqualified utilities. But in this way the justificatory relations, i.e. the reasons behind our valuations are disregarded, which is a serious omission [cf. Lumer 1998, 33-37]. From the standpoint of a stronger rationality theory we should be very interested in an empirical theory about the originally intrinsic desires behind our prospect desires, which thus would clarify more deeply the possible ways of deciding. Unfortunately, there is a severe disproportion between the extensive psychological and economic research about our

---

[9]   In the last section I distinguished between primary and secondary desirability criteria. The desirability criteria meant there were always criteria for total or prospect desirabilities. Intrinsic desirability has only now been introduced. So e.g. the primary criteria for 'total desirability' should not be confused with those for 'intrinsic desirability'.

dealing with not certain prospects and the scarce psychological research about the content of (originally) intrinsic desires.[10] The following theses are based on the latter material.[11]

4.2 Hedonic desires and weak psychological hedonism

H4   *Originally intrinsic desires:* There are three types of originally intrinsic desires: 1. hedonic desires, 2. feeling-induced desires and 3. corrected hedonic desires.

This hypothesis will now be explained and defended.

Psychological hedonism is the thesis that all sorts of feelings of the agent, namely pleasant or unpleasant bodily feelings, emotions and moods, are the objects of his originally intrinsic desires and that they are valued according to the integral of their intensity over time, where this intensity has a positive sign for pleasant feelings and a negative sign for unpleasant feelings. *Strong psychological hedonism* claims that the agent's feelings are exactly the objects of his originally intrinsic desires; *weak psychological hedonism* holds that beyond the agent's feelings there may be other objects in his originally intrinsic desirability function. Hypothesis H4 is a version of weak psychological hedonism, the first part making it a form of *hedonism* and the second and third part, which allow for non-hedonic originally intrinsic desires, making it a form of *weak* hedonism. Weak psychological hedonism seems to be a truism. Its explanatory power has been proved quite often. The real point of dispute is only if also strong hedonism is true or if there are originally intrinsic desires beyond hedonic ones and, if so, what their nature is. Therefore, only the second and third part of that thesis shall be substantiated here.

4.3 Feeling-induced intrinsic desires

All originally intrinsic desires different from hedonic desires, in particular feeling-induced desires, are always in danger of being explained away by some hedonistic reduction. So making a case for the former desires a - rather neutral - paradigm case is helpful where a hedonistic explanation is implausible.

Think of a three or four year old little girl of who has been provoked by her elder brother (ten years old), and being really furious with him she grabs his hand and bites his forearm, right above his wristwatch with all her strength. Surely our girl knows what she is doing - biting her brother's forearm - and she will have known in advance that this action will "damage" or "destroy" the aggressor, probably even that it will injure him and that he will suffer. And she has chosen well

---

[10]   An immediate problem with most of the respective psychological literature is that the expression "intrinsic (motive)" is not used in the philosophical sense. Some psychological results and empirical theories about intrinsic desires are presented in: Batson et al. 1983; Cialdini et al. 1987; Deci & Ryan 1985; Heckhausen 1989, 455-466 (as well as chs. 8-12 and 14); Maslow 1954; McClelland et al. 1989; Morillo 1992; Sansone & Harackiewicz 2000; Waterman 1990.

[11]   The following theory of originally intrinsic desires has been developed in: Lumer 2000, 428-521. In particular about feeling-induced intrinsic desires see: Lumer 1997; Lumer 2000, 477-493; about corrected hedonic desires see: Lumer 2000, 493-521.

the point where to bite him: in a place where her mouth has enough grip, and not on the wristwatch, where she would hurt herself. Damaging or destroying the aggressor seems to be the (desired) aim of that action, and there seems to be no other ulterior aim. But that the aggressor is damaged or destroyed is no feeling of the agent so that we have a non-hedonic originally intrinsic desire. Afterwards the girl will be satisfied in a crude moral way. But it seems to be too far-fetched to suppose that such a little girl already knows about such hedonic consequences of her acting out of rage; at least when she acts out of rage the first time she cannot have the empirical knowledge about these hedonic consequences, but must acquire this knowledge and perhaps some time, perhaps some years, later may even have the intrinsic hedonic aim of being morally satisfied.

The general explanation for this type of intrinsic desires runs as follows. Emotions like rage are caused by some classification, e.g.: 'This was an (unjustified) aggression by x (somebody else) against me (or somebody else one sympathizes with)', where the aggressor is neither too strong nor too insignificant. This classification implicitly is a valuation, i.e. the classificatory attributes are meant to be positive or negative. Such classifications and valuations with respect to their content are not hedonic; but they are not yet *motivational*; they are only *affective* in the sense that they cause certain emotions. So they are not the intrinsic value judgements with motivational force we are looking for. However every type of emotion is linked with another, '*satisfying*' type of emotion in the way that tokens of the first emotion seem to aim at tokens of the second, satisfying emotion. Rage has (moral) satisfaction as its satisfying match, happiness has attachment as its satisfying match, fear has relief as its satisfying match, etc. These matches again have their own affective evaluative classifications as their causes. Let us call the original type of emotion "E", the predicate used in its underlying affective valuation "V_E", the type of matching, satisfying emotion "M", and the predicate used in the underlying valuation "V_M". The mechanism behind emotion-induced intrinsic desires then is this:

H5: *Emotion-induced intrinsic desires:* If somebody is in the emotional state E and during her emotional arousal classifies some consequence or outcome o of a potential action a of herself as fulfilling the value criterion V_M of the matching emotion M (i.e. she believes: o is V_M) then she (originally) intrinsically desires outcome o in proportion to the strength of the inducing emotion of type E.

In our example the girl is in a rage (E); she considers the option of biting her brother (a) with the consequence that he will suffer (o); the matching satisfying emotion is (moral) satisfaction (M); the affective valuation criterion for causing this (moral) satisfaction is 'x is a just punishment or, a bit more primitive, a damaging or destruction of some villain or aggressor' (V_M); and the girl judges 'that my brother will suffer terribly is a just punishment of this villain' ($V_M(o)$). This judgement in a certain sense already coincides with the (positive) motivational intrinsic desire to make the brother suffer; and the strength of that desire corresponds to the intensity of our girl's rage (E).

Psychologists have tried to find out if acts of compassion might be explained roughly along these same lines [Manucia et al. 1984; replication and extension by: Cialdini et al. 1987; cf. also: Batson et al. 1983]. They tried to exclude that their subjects could aim at positive hedonic effects of

their helping. The results were, first, that emotions were necessary for helping and, second, that subjects helped independently of the fact that they could or could not expect pleasant feelings from their help. These results do not confirm the *exact* hypothesis given above. But they confirm at least the more general hypothesis that in case of acts of compassion there may be non-hedonic intrinsic desires to help other people.

Up to now we have considered only *emotionally* induced desires. There seems to be a similar phenomenon in bodily feelings and in moods as well, i.e. that these feeling states, provided they are strong enough, induce originally intrinsic desires as well. Here I can only mention this generalization of the theory of feeling-induced desires.

Feeling-induced desires seem to be an evolutionary older motivational system than the normal hedonic motivational system. And it is a much more primitive and less rational motivational system than the latter. First, emotionally induced intrinsic desires aim at rather immediate changes of the exterior situation. This change often will be beneficial to the person; otherwise this motivational system would not have survived evolution. But there is no room for changing these aims if they are not beneficial or if there are possibilities of advancing the (long-term) well-being of the person much more. Second, feeling-induced desires are bound to current feelings so that they change rapidly over time. This implies that they are not apt as a basis for long-term planning and long-term decisions. This distinguishes them sharply from hedonic desires, which are stable over time: if I know that I might suffer from hunger or anxiety one year from now I intrinsically disapprove of such feelings now, one year later or whenever. And the stability of such valuation is the basis for the fact that I can currently plan to avoid such feelings which otherwise will arise only very much later. So feeling-induced desires are not suitable as a basis for a rational desirability function. Third, feeling-induced desires cause important forms of weakness of will. We have formed a rational (in the sense just explained) judgement that *a* is (not) optimum, but when it comes to (not) doing *a* feeling-induced desires lead to a divergent emotional optimality judgment, which causes the emotionally best action.

4.4 Corrected hedonic desires

The other type of not classically hedonic but nonetheless originally intrinsic desires is *corrected hedonic desires*. These desires may come into play in cases of Nozick's experience machine [12] or in cases of improving one's feeling states by drugs (and not going the "normal" way, i.e. doing something fulfilling or rewarding) or in other cases of manipulation or alteration of one's feelings. Nozick introduced his thought experiment in order to defeat hedonism. To a certain degree he was successful. Most persons discount positive feelings that have been induced by some

---

[12]   This machine, provided by some friendly extraterrestrials, can be programmed to make you experience whatever you like if you enter the machine for ever; once having entered the machine you will not be aware that everything that you now experience is fiction so to speak. The machine is programmed in such a way that the experiences you will have in it will be significantly, though not tremendously better in hedonic terms than what you would experience in real life. Would you enter the machine? [Nozick 1989, ch. 10.]

manipulation, but not negative feelings. This means they correct their hedonic valuations. They continue to give intrinsic value (in a cool mood) only to feelings; in this respect they remain *hedonists*. But they correct the attributed desirabilities, thus becoming *corrected* hedonists. People I have asked about their reasons for this correction could not really give an answer; the corrected valuation seemed to be a question of taste. And this indicates that the valuation remains intrinsic. It is *prompted* or triggered by the reflection about the possibility of different sorts of manipulation; but the new way of valuing is not *justified* by this reflection.

H6:   *Corrected hedonism:* People who reflect on the manipulation of their feelings tend to become corrected hedonists. In their intrinsic hedonic desires they discount manipulated positive feelings according to the degree of manipulation, whereas they continue to desire negative feelings according to the assumed negative integral of the feeling's intensity over time.

Some philosophers have suggested further originally intrinsic desires, like the desires for autonomy, virtue and love.[13] But these desires may be explained rather easily in a hedonistic way, and the authors of these proposals have not provided evidence that would refute such an explanation. In addition their proposals remain *ad hoc* and isolated in the sense of not being integrated in a complete theory of intrinsic desires or motives and of desires in general - as has been sketched here.

4.5 Dynamics of intrinsic desires

Originally intrinsic valuations rely on general but usually unconscious 'criteria', i.e. they are not executed arbitrarily and differing from case to case but do show some pattern. So people value as if applying a criterion, which normally is not conscious. Such criteria for originally intrinsic valuation are inborn; people do not construct them through reflection - though they may be *triggered* by reflection, as in the case of corrected hedonic desires; and they are rather fixed. This does not imply that the criteria for intrinsic valuations are static. In what has been said so far some dynamics of this kind has already been hinted at. There are (at least) five kinds of changes of criteria for intrinsic valuations.[14]

1. Cognitive development: Criteria for originally intrinsic valuations may be changed due to some cognitive reflection. This reflection does not really provide a *reason* for the new criterion, it only *stimulates* or triggers the change - otherwise the new criterion probably would be only a criterion for prospect desirabilities. The only example for such a change found so far is the development from simple to corrected hedonism. Such changes represent real cognitive development in that the new criterion is 'based' on a more extensive cognitive background, though they are not justified but

---

13   Among others, intrinsic desires for the following objects have been proposed in the literature: autonomy [Young 1982], personal expressiveness, i.e. living in accordance with one's true self [Waterman 1990], virtue and love [Davis 1977], intellectual consistency, psychological maturity and moral rectitude [Nelson & Wilker 1975], perfection of character [Tatarkiewicz 1947/1962, ch. 24], among other things having the right contact with reality, wisdom [Nozick 1989, chs. 10; 11; 23].

14   For the dynamics of criteria for intrinsic valuation see: Lumer 2000, 205-218.

only triggered by these new cognitions. For a rational desirability theory it would be interesting to discover the perhaps final level of such cognitive developments. Corrected hedonism probably is such a final level.

2. Autonomization of prospect desirabilities: People use many particular secondary instrumentally justified criteria for prospect desirabilities. Sometimes they forget this genesis and justification and take the originally prospect desirabilities to be intrinsic desirabilities in the sense that they no longer can give a reason for their respective desires. Whereas cognitive development of criteria for intrinsic valuation is a cognitive progress, such autonomization of originally prospect desirabilities to consciously intrinsic desirabilities is a sort of step back because it is based on forgetting.

3. Dissolution of autonomized desirabilities: However this regression may be caught up with. If people learn about the genesis of their consciously intrinsic desirability criterion they tend to no longer use it as such. Either they keep on holding the old (or a better) justification for this criterion to be valid; then they continue to use it, but once again (as originally) as a criterion for prospect desirabilities. Or they find the old justification to be fallacious and see no better substitute for it; then they tend to stop using the criterion.

4. Repression of originally intrinsic desirability criteria: If some sorts of originally intrinsically good objects (e.g. sexual pleasure or pleasure resulting from aggressive behaviour, which lead to punishment) over a longer period generally have strong negative consequences people may repress their respective intrinsic desires: they value the respective objects as prospectively negative and 'forget' their positive intrinsic component. Such repression is a cognitive step back, too, because it is based on forgetting and because it tends to overrate the negativity of the respective objects (because their positive component is no longer available).

5. Recovering the repressed: However this kind of regression may be caught up with, too, by restoring the repressed intrinsic valuation to people's consciousness.

H7: Dynamics of intrinsic desires: Intrinsic desirability functions can change as a consequence of the following processes: cognitive development, autonomization of prospect desirabilities, dissolution of autonomized desirabilities, repression of originally intrinsic desirability criteria and recovering the repressed.

An important consequence of these mechanisms of changing intrinsic desirability function is that more information tends to lead to a stable state with maximally cognitively developed intrinsic desirability functions, where all perhaps at some time autonomized desirability criteria are dissolved and all perhaps at some time repressed originally intrinsic desirability criteria are reinstalled.

5. Motives for moral action

Completing all the details of the picture outlined so far is not possible here. But some further details of particular interest for practical philosophy shall be sketched in this section. They regard motives for moral action.

Actions in accordance with the requirements of morals may be motivated in rather arbitrary ways. But there are some motives *substantially* connected to acting morally, in particular the following.[15]

H8: Moral motives and motives for moral action: The following motives are substantially connected to acting morally.

H8.1. Self-interested cooperation: Informed prudential self-interest in reciprocal cooperation can lead to respecting another person's interests.

H8.2. Sympathy: Sympathy is an emotion and not a motive. But it can lead to moral action in two different ways. 1. Being an emotion, sympathy can induce an intrinsic desire to improve another person's situation. However such desires are not temporally stable (cf. above, sect. 4.3). 2. Pity is a negative, unpleasant emotion; it has a positive, pleasant counterpart, a sort of satisfaction about someone else's positive well-being. Being pleasant or unpleasant, sympathy falls in the domain of hedonism, and one can try to optimize one's sympathy: one can change the world in such a way as to have more occasions for feeling positive sympathy and fewer occasions for feeling negative sympathy.

H8.3. In-group solidarity: A certain form of solidarity, which may be called "in-group solidarity", is based on one's identification with a certain group. This leads to augmenting the salience of and the interest in the fate of in-group members. This in turn can focus sympathy, which in principle is universal, as to be dedicated to a stronger degree to the proper in-group. The most ambitious and universal identification is the identification with mankind; less ambitious identifications can e.g. be nationalistic or have the proletariat or all oppressed people as their scope.

H8.4. Respect: One form of respect are emotions out of a group of emotions like admiration, reverence, fascination, being deeply moved, which are caused by an appraising classification of a certain object, a person, animal, nature, artefact etc., as highly developed, elaborated or ingenious etc. and fragile. Like sympathy, respect can lead to moral action in two ways. 1. These emotions can induce a motivation out of respect, namely a positive intrinsic desire of the respective object's existence, which in turn may lead to appropriate action like protecting this object. 2. On the other hand one may seek to optimize respect emotions by creating occasions for positive respect emotions by means of protecting and supporting the object.

H8.5. Self-transcendent motives: There is a group of self-transcendent emotions like love, affect, liking, creative self-expansion and pride in the proper community or culture. All these self-transcendent emotions arise from a deep connection to a rather particular object. And they can induce motivation to protect the respective object and to make it flourish.

H8.6. Moral motives: Finally, there are moral motives in the narrow sense, that is motives originating from some moral judgement and translating it, so to speak, into moral practice. 1. There are, once again, emotion-induced moral motives, originating from feelings of guilt or indignation or outrage and which motivate to punishing or, more rarely, to repairing. 2. On the other hand there is a hedonist version of moral motives, namely to raise one's moral self-esteem (and to prevent its

---

[15]   For a more extensive analysis of motives for moral action see: Lumer 2002.

being lowered) by acting morally and then positively appraising one's deeds or character. Positive self-esteem after all is (or, in the mere cognitive meaning of "self-esteem", results in) a pleasant feeling, and negative self-esteem is an unpleasant one. Therefore, we can be hedonistically motivated to optimize our moral self-esteem by acting morally.

6. Why normative theories of practical reason presuppose empirical theories of practical reasons

After having sketched the empirical theory of practical reasons, we can now draw some normative consequences. The central thesis underlying the second part of this paper is: Normative theories of practical reason, in particular the theory of practical rationality and ethics, presuppose empirical theories of practical reason in the sense that the former must be based on rather strong empirical information about the psychological laws governing our deliberations and decisions.

This thesis holds for the following reasons. The theory of practical rationality is normative in the broad sense; more precisely it makes *recommendations* for deciding and acting rationally. Ethics on the other hand is (in part) normative in the strict sense, but it is also normative in this broad sense, it makes recommendations for deciding and acting morally. Now reasonable recommendations must fulfil the following adequacy conditions.

A1: Realizability: The recommendation must be *realizable*.

A2: Selectivity: The recommended course may not be such as to be followed necessarily anyway (e.g. due to some law of nature); i.e. the recommended course has to be *selected* from a possible range of open courses. Otherwise the recommendation would be void.

A3: Goodness: The recommended course must use the open spectrum of different ways of deciding and courses of action in a *good* manner so that the recommendation is appealing to the addressee. Otherwise there would be no reason to follow the recommendation. Expressed a bit differently: the proposals must be designed in such a way as to make people, with the help of some argument, inclined to follow these proposals.

These are only *necessary* adequacy conditions implying something about the empirical underpinnings of normative theories of practical reasons. Here is not the place for developing sufficient adequacy conditions.

Now fulfilling these adequacy conditions requires precise empirical information about our ways of deciding. The argument for this claim is straightforward. Ad A1: The recommendations do not regard (executive) action only but also our way of deciding. Therefore, we must not only be able to *act* in the required way (where "to be able to do *a*" is meant in the conditional sense, i.e. as: if I intend to *a* I will do *a*); we must also be able to follow the suggested decision rules (this holds for the theory of practical rationality) and to arrive at the required intention (this holds for normative ethics in the narrow sense). And for checking if the addressees of our theories are able to do so, and for bringing it about that they really do so we must know how people decide and, in particular, how their ways of deciding react on which input. Ad A2: For fulfilling the selectivity

condition we must know which features of people's ways of deciding are necessary and where there are options for deciding differently. Ad A3: For making good recommendations for our ways of deciding and for the content of our intentions we must at least know which options people have for deciding differently and what the consequences of these options are. - Note that these arguments do not depend on an internalist or externalist conception of normative practical reasons.

The first parts of each of the following two sections give some examples of how normative theories of practical reason fail because they are not at all or not sufficiently based on empirical theories of practical reasons and therefore violate these adequacy conditions. The second parts of the following two sections, on the other hand, show how, on the basis of the sketched empirical decision theory, a satisfying ethics and theory of practical rationality that satisfies the adequacy conditions can be developed.

7. Consequences for the theory of practical rationality

A first implication of the above sketched empirical theory for non-moral practical philosophy is that it tends to exclude libertarian views of free will in that respect that it shows how decisions are subjected to empirical laws (H1-H8).[16] On the other hand this theory opens the way to a form of freedom consisting in deciding rationally and maximizing the individual good. Another interesting implication is that, according to H1, optimality beliefs simultaneously have *two* directions of fit: as beliefs they try to fit to the world, that is, to be true and to identify the really best action (in subjective terms, though); as intentions they aim at making the world fit to the intended course of action, that is, to the action held to be optimum. A third important implication regards the question of instrumentalism: Can actions be rationally criticized only with respect to the means used or also with respect to their (ultimate) ends? On the one hand, the originally intrinsic desirability functions described in subsections 4.2-4.4 and thus the possible originally ultimate ends are fixed. On the other hand, the above sketched dynamics of intrinsic desires (subsect. 4.5) leaves some room for rationally criticizing actual intrinsic desires: It is possible that actual intrinsic desires are not originally intrinsic, not cognitively maximally developed or temporally unstable; in all these cases they are unstable with respect to acquiring new information. A rational theory of ultimate ends that tries to fulfil adequacy condition A1 (realizability) can use this room but is bound to its limits. In the rest of this section some consequences for the theory of practical rationality will be developed more systematically.

7.1 Criticisms of rational decision theory and full information accounts of rationality

Rational decision theory on the lines of a von Neumann / Morgenstern-like axiomatization does not fulfil the adequacy conditions A1-A3. In particular its recommendations are not realizable

---

16   Of course, the hypotheses sketched above, as usually in psychology, at best report very high frequencies and thus cannot exclude undetermined decisions altogether.

(cf. A1). Nobody fulfils the axioms of these theories [Kahneman & Tversky 1979, 263-273; Camerer 1995, 622-626; 644-670; cf. above, H2]. But fulfilling the axioms is a necessary condition for subjective utilities to be defined. Therefore, the subjective (expected) utilities for the respective person are not defined, and consequently it is impossible to calculate expected utilities of the options at hand. In addition, even if one could calculate them in principle, in most cases it would be practically impossible or irrational to calculate them because this would be too time-consuming (cf. A3). This means rational decision theory as such does not tackle the problem of deciding how to decide (cf. H2 and H3) and does not take the problem of scarce decision resources behind H2 seriously enough. Finally, the theory is pragmatically useless. It states only what a rational system of preferences and decisions looks like. But it does not give us procedural rules on how to arrive at that ideal final state. Even in this pragmatic sense the recommendations of that theory are not realizable (cf. A1).

But many philosophers as well as some economists too have criticized rational decision theory from a quite different point of view that aims at completing rather than replacing it. They say that rational decision theory's standards of rationality are too low. In particular, they say that it is irrational to accept people's unfiltered preferences uncritically, even if they are coherent, as a basis for defining prudential desirabilities. After all these preferences may be based on false information [Harsanyi 1982, 55 f.]. This criticism implies that rational decision theory does not recommend the best way of deciding, thereby violating again A3.

The full information approach to practical rationality seems to be the most obvious and insurmountably rational reply to the latter criticism.[17] Roughly, this approach holds that a decision is rational and represents the choice of the best option if it would be the result of a decision made with complete (relevant) and vividly represented information. Full information approaches to practical rationality and desirability face severe problems, too. First, we never have full information and even if we had we could not represent it all vividly at the same time. Therefore, we can never check empirically which of two options is prudentially better (according to full information theory). And the approach does not provide the means for finding out which alternative would be preferred. Thus A1 is violated.[18] Second, in a certain sense full information approaches are rationally insurmountable because the person, at least hypothetically, always knows everything that is relevant. However this is exactly the problem. If the person really had all the *empirical* information she still would not know what to do with it, which way of deciding to follow. So one has to add the

---

[17]   Richard Brandt [1979, 10-13; 70-88; 110-129] has developed the most important proposal of this kind. Another full information approach is e.g.: Griffin 1986.

[18]   To be fair it must be recognized that Brandt has undertaken particular efforts, more than any other theoretician of practical rationality in the last decades, to insert psychological information into his rational model for finding out what the result of a fully informed decision would be [Brandt, Good 24-148]. However, first, due to its very nature, the late behaviouristic psychology he used is more than imprecise and cannot capture and explain what cognitive processes lead to our decisions. And second, these psychological laws can tell us only what the effect of various stimuli would be but nothing about the response on full and vividly imagined information. So Brandt's assumptions about the results of rational decisions are rather speculative.

information about the best way of deciding to the full empirical information. But this would be precisely the information that the theory of practical rationality should provide. So the full information approach presupposes that the person hypothetically has solved the general problems of rationality theory, which even the theoretician himself has not been able to solve so far. This is asking a bit too much. Therefore, the realizability condition (A1) is violated again. So even the full information approach (like rational decision theory) is pragmatically useless (cf. A3) since it does not give procedural recommendations on to how to arrive at rational valuations. And one reason for this is that the proposals of this approach *refer* to the empirical outcome of certain decisions but do not make sufficient *use* of the empirical information about our ways of deciding.

7.2 Reconstructive prudentialism

An approach that avoids these problems and nonetheless reaches far beyond the rationality level of rational decision theory is a reconstructive desirability theory, which relies heavily on empirical information about our decision making. The basic idea of this approach is to inquire the empirically possible ways of deciding and of establishing desirabilities, and to recommend that way as the way for calculating prudential desirabilities that remains stable in view of new knowledge (and perhaps fulfils some further conditions). More precisely, this approach gets along with the following steps. First, it has to be reconstructed how people really decide, or more exactly, which possible ways of deliberation and deciding exist and under what conditions they are taken. Second, it has to be found out which of these ways is stable in view of new knowledge and which fulfils further requirements like being stable in other respects. Exactly this way then defines prudential 'desirability'. (Thus all three adequacy conditions, A1-A3, are fulfilled.) So this definition has taken over from the agents only the *way* of deciding and only the *wisest* way of deciding (wisest in the sense of relying on all relevant structural information so that stability is reached) and thus the purely conative part of their deliberation. The *empirical information* necessary for finally establishing the (real) desirability of a given object must be objectively true and can be provided by whoever has it; so the empirical information is inserted from the outside. In this way among others the overflow of information in full information approaches is avoided (cf. A1). Third, based on this prudential definition of 'desirability', more convenient (i.e. easier to apply or requiring less information) secondary criteria for 'desirability' can be developed, which then should be used in rational decision (cf. A3).

Obviously, this prudential and reconstructive approach requires very strong empirical information about our ways of deliberation, that is, an empirical theory of practical reasons. The respective theory presented above (sects. 2-5) roughly leads to a desirability theory of the following type [detailed exposition: Lumer 2000, 241-427; 521-548]. 1. The gross structure is anthropologically fixed and consists of three parts: intrinsic desirabilities (H4-H7), prospect desirabilities, defined in terms of intrinsic desirabilities (H2-H3), and optimization (H1). 2. If we apply the criterion of stability in view of further information and if we consider what has been said about the dynamics of intrinsic desirability criteria (H7) only *original criteria* of 'intrinsic

desirability' are eligible as prudential criteria of 'intrinsic desirability' and among these only the cognitively most developed ones (if there is any cognitive development at all). The only empirical criteria fulfilling these conditions are corrected hedonism and the criteria inherent in feeling-induced intrinsic desires. The latter desires strongly depend on the underlying feelings (H5), however, and thus are completely unstable over time and do not permit any long-term planning. Therefore, for formal reasons of rationality the latter group should also be excluded as basis of prudential intrinsic desirabilities. So we end up with corrected hedonism as the prudential criterion of intrinsic desirabilities (cf. H6). 3. The fundamental definitions of 'total desirability' and 'prospect desirability' must follow the lines set out in the adequacy condition hypothesis about fundamental prospect desirability concepts (H3). 4. An empirical theory of deliberation, finally, is heuristically very rich in providing a huge wealth of secondary criteria of prospect desirability, which then should be scrutinized as to if and under which conditions their application is optimum according to the fundamental criteria.

This theory of prudential desirability fulfils the three adequacy conditions A1-A3. One might suspect that it is impossible to practically neglect the criteria of 'intrinsic desirability' that have been discarded as irrational, so that A1, realizability, would be violated. But this is not true. Autonomized prospect desirabilities, which have become conscious intrinsic desirabilities, can be dissolved by confronting people with the history of these intrinsic desirabilities. Simple hedonism may be overcome by confronting people with thought experiments of the experience machine type. And feeling-induced desirabilities can be made ineffective by delaying one's decision until the respective feelings have cooled off a bit. (This actually is what folk wisdom advises: Do not decide in the heat of passion because you may regret your decision later!) It is rather obvious that the recommendations of reconstructive desirability theory are not followed anyway (A2) because it proposes to follow precisely one of the many possible ways of deciding. It may be difficult to prove that the recommendations of this theory are good (A3) and as good as possible among the recommendations of competing theories that also fulfil conditions A1 and A2 because such theories just *define* 'good' etc. Nonetheless the suggested theory has some clear advantages. First, filtering out ways of deciding that are unstable in view of further knowledge or during the course of time prevents false investments, i.e. investments in projects whose success later is regarded to be vain or even harmful. Second, stability in view of further information guarantees something like wisdom and seeks to satisfy ideals of enlightenment.

8. Consequences for ethics

What are the consequences of the above sketched empirical theory of practical reasons for ethics? A straightforward example for critical consequences of this theory is utilitarianism. The standard interpretation of utilitarianism obliges us to always maximize moral desirability. It does not take into account that, aside from their moral motives people have much stronger self-interested motives (cf. H4-H6), which make it impossible for them to obey that obligation. So condition A1

(realizability) is violated. In order to fulfil adequacy condition A1 moral obligation has to be designed on the basis of information about the strength of our moral motivation and about other motives that might be apt to support the fulfilment of moral obligations. Another consequence of the empirical theory outlined above may be that the axiological character of practical reasons assumed by it (cf. H1-H3) strongly favours an axiological, e.g. welfarist, ethics as opposed to deontological or virtue ethics. Deontologism e.g. seems to presuppose a decision psychology of "acting according to principles / subjective maxims", where the will is always acting according to the representation of a law [cf. Kant, GMS BA 15; 36; 63]. Of course, humans can do this - but always on the basis of respective desires (cf. H2-H4). So, given humans' axiological character, deontologism will always have difficulties in making us motivationally accept deontologically justified duties (cf. A1).[19]

These criticisms are rather simple. Consider a more complex example, foundational externalism.

8.1 Critique of foundational externalism

Most ethicists accept what, introducing some distinction here, may be called "*effective internalism*", i.e. the thesis that a sound moral justification must motivate rational or prudent persons to follow that moral to at least some degree - which does not exclude that in the end there may be much stronger motives to the contrary. (Some ethicists, like Brink [1989] or Schaber [1997], do not accept effective internalism. But if their ethics do not fulfil the requirements of effective internalism, i.e. if they are not motivating, their ethics have little practical sense; even as theorists we can ignore them because people in their practical decisions will do so as well.) Effective internalism is an implication of adequacy condition A3 (goodness): the ethics proposed must be practically appealing. *Foundational internalism* is a stronger thesis: it accepts effective internalism and claims that for fulfilling the motivational requirement moral justifications have to rely on motives that precede these justifications [Williams 1979]. (*Foundational externalism* on the other hand claims that the motivational requirements of effective internalism can be fulfilled independently of such pre-existing motives. In a certain sense the justification should *create* the right motivation.)

Kant, according to these definitions, is an effective internalist because he not only wants people to follow the morality he justifies but he also wants them to follow it for the right motives [cf. e.g. Kant, GMS BA X; 26]. At the same time he is a foundational externalist. His basic idea is that pure reason has to establish what is morally right independently of any motivational inclinations of men and, therefore, independently of empirical theories as introduced above. Only in a second step the insights of pure reason shall motivate to action, thus fulfilling the requirements of effective internalism. [Kant, GMS BA 33; BA 36 f.; BA 63 f.; KpV A 56-58; A 126-128.] So Kant,

---

19   This argument and several other respects in which axiological ethics fit better to human decision psychology are elaborated in: Lumer 2004b.

like today's foundationally externalist ethicist, denies the relevance of empirical theories of practical reasons for the foundational tasks of ethics altogether. However, for fulfilling the requirements of effective internalism empirical information like the one introduced above is indispensable because motivating for action is a causal and, according to a Humean type of causality, contingent empirical relation.

In other words: Kant's original project, to find out moral laws by pure reason and then to let the respective insights alone determine action, is even *analytically* impossible (cf. A1). This holds because to determine action or to motivate to act is a causal relation, and causal relations depend on respective empirical laws. According to the idea of pure practical reason, very specific cognitions about an action, say that this action is *F*, should cause the execution of that action, whereas different cognitions, that the action is *G*, *H* etc., should not have this effect. But if this is to happen then there have to be psychic mechanisms, describable by means of psychological laws, bringing about that exactly cognitions of this kind, namely that an action is *F*, lead to the action's execution. Now, which of the many possible qualities *F*, *G*, *H* etc. of an action recognizable by pure reason exactly is relevant for executing the respective action is fixed by those *mechanisms* - and not by pure reason. Pure reason cannot decree the practical relevance of its cognitions. [Lumer 2002/03, 269-271.][20]

What has been said so far does not exclude a more down to earth version of foundational externalism in ethics, though. There may be motivational mechanisms that translate moral judgements into some respective motivation rather independently of the specific content of the moral judgement. According to this hypothesis, a moral conception may be justified independently of motivating reasons; and if someone is convinced of this conception and of the truth of certain moral judgements about an action then that mechanism induces some motivation to act accordingly. Differences with respect to what Kant intended would be, first, that the whole construction would depend on the 'contingent' existence of such an empirical mechanism in humans, which also would determine the kind of cognition to which we would be motivationally responsive, and, second, that this mechanism also would limit to a certain degree the spectrum of moral conceptions that might be motivationally supported. But such a mechanism coincides exactly with what above (H8.6) has been called "moral motives" (in the narrow sense). The prudentially (and also in moral practice) most important moral motive is raising one's moral self-esteem (H8.6.2). However, whether seeking positive self-esteem, finally, is a sufficient basis for foundational externalism in ethics, still depends on the moral neutrality of this mechanism. Does this mechanism support (within certain limits) whatever moral conceptions, so that pure reason would be free to find out the right morals? The answer to this question now depends on what moral conceptions we can rationally adopt in such a manner that the adoption leads to influencing our self-esteem in the right way. This is a question regarding the psychology of *moral emotions*, which is beyond the scope of this paper. But

---

[20]   Similar criticisms can be advanced against all versions of value objectivism and moral realism whose values are not bound to psychic mechanisms and which nonetheless hold that recognition of the value object generates motivation to act in favour of these objects.

the fact that emotions, too, are caused by very specific and anthropologically fixed judgements makes it unlikely that our moral self-esteem should be open to whatever moral conceptions, as foundational externalism requires (cf. A1).[21]

8.2. The basis of foundational internalism

Foundationally internalist approaches in ethics assume that, for providing understandable reasons and for being able to motivate, justifications of morals must be based on (in a certain way) pre-existing rational motives, which later on shall also be the reasons for following the justified morals. The above sketched theory of motives for moral action (H8) has presented the candidates that may serve as such a basis. But moral justification and motivation should be practically rational, which includes stability in view of new knowledge. Not all motives for moral action listed in H8 fulfil this requirement and thus are not suitable as the basis of an internalist justification of morals.[22] Sympathy-induced motives (H8.2.1) and motives induced by respect (H8.4.1) as well as emotion-induced moral motives (H8.6.1) are feeling-induced motives, hence unstable over time. Self-transcendent motives (H8.5) may function as the basis of an individual morality, but they are too idiosyncratic for being the basis of a social and universalistic morality. All moral motives in the narrow sense (H8.6), including those not induced by emotions, finally, are not apt as such a basis either because they already originate from moral judgements; therefore, a justification strategy based on these motives would lead to a vicious circle. All the motives mentioned so far play more or less important roles as moral motivation *amplifiers* but they are not apt to provide the initial moral "*signal*". The remaining motives for moral action of the list in H8.6, on the other hand, can serve as justifying motives in an internalist foundation of morals, and have mostly been used as such in the history of philosophy. Based on a prudential and informed self-interest in cooperation (H8.1) a weak business morality, without altruism and redistribution of utilities, can be justified - as it has been done e.g. by Gauthier [1986]. Sympathy optimization (H8.2.2) may be the basis of a universal morality of benevolence as British moral sentimentalists intended it, whereas the *locus classicus* for a sympathy-based morality is Schopenhauer. In-group solidarity (H8.3) is only a mechanism for altering the extent of sympathy, in particular increasing sympathetic motivation in favour of the members of one's in-group. This feature is apt to justify hierarchical pathocentrism for example. Optimizing respect emotions (H8.4.2), finally, may be the basis for justifying a prohibitive morality of respect, such as the principle 'Do not harm anybody!' or an ethics of conservation of nature and culture.

---

[21]   The ontogenetic beginning of morality is heteronomous in that morality is taken over from the socialization agents and thus justified by authority. But later, in particular beginning with adolescence, people may not only systematize their morals and question the parts that they find unjustified, they also bring in new autonomous sources of morals, which then can become and remain stable bases of their morals. These autonomous sources of morals seem to be exactly those motives, however, that *internalist* justifications of morals are based upon: rational self-interest in cooperation, sympathy and respect. [Lumer 2002, sect. 7; Lumer 2002/03, sect. 8.1.]

[22]   For a more extensive justification of the following assessments of the motives for moral action see: Lumer 2002, 171-182.

Because all the latter motives can be used as a basis for justifying morals, the question of how to integrate all these motives into one morality or how to exclude some of them arises. But this question is beyond the scope of this paper. However, the empirical theory of practical reasons sketched here could already be shown to be useful for an internalist approach to ethics in individuating the motives that are suitable as a basis for such a foundational task. Working out this ethics up to precise moral criteria then requires much more scrutiny of the exact features of the respective motives. For example the exact limits of rational self-interest in cooperation will also define the limits of the business morality, or the exact features of our sympathy will also define the degree and distribution of benevolence in a sympathy based ethics of benevolence etc.[23] This means, in practical philosophy we need still much more of an empirical theory of practical reasons than what could be outlined here.[24]

References

Batson, C. Daniel; Karen O'Quinn; Jim Fultz; Mary Vanderplas; Alice M. Isen (1983): Influence of Self-Reported Distress and Empathy on Egoistic Versus Altruistic Motivation to Help. In: Journal of Personality and Social Psychology 45. Pp. 706-718.

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.

Brink, David Owen (1989): Moral Realism and the Foundation of Ethics. Cambridge: Cambridge U.P. xii; 340 pp.

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.

Cialdini, Robert, B.; Mark Schaller; Donald Houlihan; Kevin Arps; Jim Fultz; Arthur L. Beaman (1987): Empathy-Based Helping. Is It Selflessly or Selfishly Motivated? In: Journal of Personality and Social Psychology 52. Pp. 749-758.

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.

Dancy, Jonathan (2000): Practical Reality. Oxford: Oxford U.P. xii; 187 pp.

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.

Davis, William H. (1977): Transcendent Needs. In: Philosophy Today 21. Pp. 184-193.

Deci, Edward L.; Richard M. Ryan (1985): Intrinsic Motivation and Self-Determination in Human Behavior. New York; London: Plenum Press. XV; 371 pp.

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

Gauthier, David (1986): Morals by Agreement. Oxford: Clarendon. ix; 367 pp.

---

[23] An attempt to establish the exact features of our sympathy and to draw (prioritarian) ethical consequences from them is: Lumer 2000, 589-632.

[24] I would like to thank several people for discussions of this paper, in particular: Robert Audi, Michael Bratman, Carl Ginet, Hugh McCann and Alfred Mele.

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

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

Griffin, James (1986): Well-being. Its meaning, measurement, and moral importance. Oxford: Clarendon. xii; 412 pp.

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.

Harsanyi, John C. (1982/<1977>): Morality and the Theory of Rational Behaviour. In: Amartya Sen; Bernard Williams (eds.): Utilitarianism and Beyond. Cambridge: Cambridge U.P. 1982. Pp. 39-62.

Heckhausen, Heinz (1989): Motivation und Handeln. 2., völlig überarb. u. erg. Aufl. Berlin [etc.]: Springer.

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

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 (1998): Which Preferences Shall Be the Basis of Rational Decision? In: Christoph Fehige; Ulla Wessels (eds.): Preferences. Berlin; New York: de Gruyter. Pp. 33-56.

Lumer, Christoph (2000): Rationaler Altruismus. Eine prudentielle Theorie der Rationalität und des Altruismus. Osnabrück: Universitätsverlag Rasch. 652 pp.

Lumer, Christoph (2002): Motive zu moralischem Handeln. In: Analyse & Kritik 24. Pp. 163-188.

Lumer, Christoph (2002/03): Kantischer Externalismus und Motive zu moralischem Handeln. In: Conceptus 35 (2002/03). Pp. 263-286. - Abridged Version: www.gap-im-netz.de/gap4Konf/Proceedings4/Proc.htm.

Lumer, Christoph (2004a): Desires and Explanatory Reasons. In: Roland Bluhm; Christian Nimtz (eds.): Selected Papers Contributed to the Sections of GAP.5, Fifth International Congress of the Society for Analytical Philosophy, Bielefeld, 22-26 September 2003. (CD-ROM.) Paderborn: Mentis. Pp. 704-714.

Lumer, Christoph (2004b): Vom Primat der Werte - Wertethik versus Pflicht- und Tugendethik. In: Reinhold Mokrosch; Elk Franke (eds.): Wertethik und Werterziehung. Göttingen: Vandenhoeck & Ruprecht unipress 2004. Pp. 39-61.

Lumer, Christoph (2005): Intentions Are Optimality Beliefs - but Optimizing what? In: Erkenntnis 62. Pp. 235-262.

Manucia, G. K.; D. J. Baumann; R. B. Cialdini (1984): Mood Influences on Helping. Direct Effects or Side Effects? In: Journal of Personality and Social Psychology 46. Pp. 357-364.

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

McClelland, David; Richard Koestner; Joel Weinberger (1989): How do self-attributed and implicit motives differ? In: Psychological Review 96. Pp. 690-702.

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.

Morillo, Carolyn R. (1992): Reward Event Systems. Reconceptualizing the Explanatory Roles of Motivation Desire and Pleasure. In: Philosophical Psychology 5. Pp. 7-32.

Nagel, Thomas (1970): The Possibility of Altruism. Oxford: Clarendon. 7°; 148 pp.

Nelson, Jack; David Wilker (1975): Pleasure and the intrinsically desired. In: Analysis 35. Pp. 152-159.

Nozick, Robert (1989): The Examined Life. Philosophical Meditations. New York [etc.]: Simon & Schuster. 309 pp.

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

Sansone, C.; J. M. Harackiewicz (2000) (eds.): Intrinsic Motivation: Controversies and New Directions. CA: Academic Press.

Schaber, Peter (1997): Moralischer Realismus. Freiburg; München: Alber. 406 pp.

Searle, John R. (2001): Rationality in Action. Cambridge, Mass.; London: MIT Press. xvi; 303 pp.

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

Smith, Michael (1987): The Humean Theory of Motivation. In: Mind 96. Pp. 36-61.

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.

Tatarkiewicz, Wladyslaw (1947/1962): Analysis of happiness. Transl. from the Polish by Edward Rothert (chap. 1-11). Danuta Zieliuskn (chap. 12-25). The Hague: Nijhoff [etc.] 1976. xii; 356 Pp.

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

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.

Waterman, Alan S. (1990): Personal Expressiveness. Philosophical and Psychological Foundations. In: The Journal of Mind and Behavior 11. Pp. 47-74.

Williams, Bernard (1979): Internal and External Reasons. In: Ross Harrsion (ed.): Rational Action. Studies in philosophy and social science. Cambridge; London: Cambridge U.P. Pp. 17-28. - Reprinted in: Id.: Moral Luck. Philosophical Papers 1973-1980. Cambridge: Cambridge U.P. 1981. Pp. 101-113.

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.

Young, Robert (1982): The Value of Autonomy. In: Philosophical Quarterly 32. Pp. 35-44.