How to Interpret Human Actions (Including Moral Actions)

Christoph Lumer (University of Siena)

Appeared in: Peter Machamer; Gereon Wolters (eds.): Interpretation. Ways of Thinking about the Sciences and the Arts. Pittsburgh, PA.: Pittsburgh U.P. 2010. Pp. 130-157.

Abstract: This paper presents two methods for interpreting human actions or their results with the aim of finding the underlying intention, the goal intention or the more comprehensive intention including the subject's considerations about various advantages of the chosen option as well as about its alternatives. The first interpretation method is simple and uses statistical syllogisms. The second method consists of an inference to the best explanation; it is much more elaborate but it provides more exact and detailed information. This method requires covering laws that explain the formation of intentions. Such laws are sketched too. They also cover moral actions so that the presented interpretation methods can also be used for interpreting moral actions.

1. Aim and Structure of this Article

In this article an instrumentalistic conception of action interpretation will be developed. This conception shall be suitable for interpreting moral actions as well as other actions. The approach's instrumentalism consists in the fact that interpretations here are conceived as means for fulfilling a certain function, in particular for providing a certain type of information about the action. After these preliminary remarks it will be explained which kind of information we expect from action interpretations (sect. 2). In the subsequent section it will be discussed which model of action and action interpretation in principle could provide this information (sect. 3). In the second part of the paper some simple methods (sect. 4) and a complex method of action interpretation will be explained (sects. 5-6). The final section is dedicated to the question of how moral actions function and whether the methods explained so far are suitable for interpreting them (sect. 7).

2. The Structure of Intentions and the Cognitive Aims of Action Interpretation

For a better understanding of what kind of information the interpretation of actions should and can provide, it is useful to consider the process preceding an action. The most widely held, namely the causal, conception of action says that an *action* is a behaviour that is caused by an intention, where the action normally corresponds to the intention. (In times of doubts about mental causation, it must be specified that here "the intention *causes* the behaviour" always means that the physical supervenience basis of the intention causes the behaviour. However, as the attentive reader will notice later on, the model of interpreting actions proposed here presupposes only an unspecific causal relation between a decision and the following action and is fairly neutral with respect to how this kind of causality is to be interpreted; monism, dualism - even Cartesian -, epiphenomenalism (cf. the preceding sentence) are all compatible with the model.) There are *automatic actions* by which we react, without further consideration, to some triggering signal; in this case the intention is not the triggering cause but still the structuring cause, formed some time ago, perhaps so long ago that the intention is no longer consciously accessible - think e.g. of the general intention to brake one's car in front of a red traffic light, which goes back at least to the time of driving lessons. On the other hand, there are actions preceded by *deliberation*, i.e. a process of practical reflection with the function of determining what to do. An *intention*, in a first approximation, is the substantial result of such a deliberation.

More precisely however, several concepts of 'intention' have to be distinguished.[1] Sometimes during a deliberation we reconsider and retract some consideration, such as a belief or a desire, already undertaken during the deliberation and we may replace it with a new idea. The *comprehensive intention*, then, consists of (i) all the thoughts (beliefs, desires) that were part of the deliberation, contributed to the final decision and were not retracted later during the deliberation, and it consists of (ii) the subjective justifications (held at the time of the deliberation) of such thoughts. A necessary part of the comprehensive intention is the *implementation intention*; this is the comprehensive intention's final and core element that determines, in a way that is understandable to the executive system, which behaviour shall be executed. Often the comprehensive intention also comprises a *goal intention*; this is a first or intermediate determination of what to do in terms of fixing some (mostly desirable) state that should be caused, or more generally: brought about, by the behaviour to be chosen: 'I want / intend to do something which brings about that *p*' - e.g. 'I want to get Smith's new paper', i.e. 'I want to do something that will bring about my possession of Smith's new paper'. Goal intentions, as opposed to implementation intentions, usually are not understandable to the executive system because the intended behaviour is only described but not *identified*; this can be seen from the existential quantification in the behaviour description: 'to do *something* which brings about that *p*'. The remaining phase of the deliberation, after determining the goal intention, serves to identify such a behaviour, i.e. to find a suitable implementation intention. Apart from the goal intention and the implementation intention, a comprehensive intention can include ideas about possible options, beliefs about the circumstances and the possible implications of the options, in particular their consequences, the probability of such implications, beliefs or desires with respect to the utility of such implications, integrations of several beliefs and desires of an option into an all considered judgement or desire regarding this option etc.

My general thesis about *the function of action interpretations is that they serve to identify all or parts of the comprehensive intention underlying an action*. (In very rare cases such interpretations also try to identify the exact course of the deliberation including the pieces later withdrawn.) This thesis is to be understood in two ways. First, it shall be a description of a common practice of action interpretation. Second, it shall be a weak normative thesis in the following sense and sustained by the following argument: As the subsequent considerations will show, we often are interested in information about other people's comprehensive intentions - which frequently are not easy to discover -; therefore, we need an instrument to ascertain intentions. The conception of action interpretation to be developed here is such an instrument, which in particular exploits the already available knowledge about the behaviour or its consequences. So, if the proposed conception of action interpretation can stand up to these expectations this would show the usefulness of this conception. This usefulness does not exclude the possibility that there are other equally useful conceptions of interpreting actions - though, so far I do not see any other important useful conception. And conversely, the possible existence of such further conceptions would not diminish the value of the instrumentalist conception of action interpretation sketched here.

Why, in which situations and on what basis do we want to know what about the intention of other people?

The starting case is that we know the other person's behaviour and want to know the underlying intention but have no relatively direct access to it, e.g. asking this person. 1. Our desire to know may be simply for sympathetic reasons - we want to understand how our child or partner or another person close to us thinks and feels - or out of curiosity. 2. Sometimes we want to find out the semantic meaning of an ambiguous or incomplete oral text (a speech, an enigmatic allusion of a rival, boss, beloved etc., an inarticulate remark in a noisy environment ...) by reconstructing the speaker's or writer's respective intention. 3. In other cases we want to know many details of the comprehensive intention for *judging* the action and the person from a moral, juridical or prudential point of view. Was the killing a murder or a hunting accident, i.e. did the agent foresee the fatal consequence? Was it his goal? Did the agent, by lending his car with the broken brakes to his rival, provoke the accident intentionally, knowingly, recklessly or negligently, i.e. did the agent have the goal intention to provoke such an accident, did he only know about the broken brakes and the probable consequence, did he know about this possibility, however attributing only a small probability to it, did he think of possible problems of the brakes but not of the consequences etc.? Did the agent kill his rival in the heat of passion because he had been provoked, or had he been waiting for such an occasion to have an excuse to kill him? The agent could have helped the woman with the heart-attack by giving her his own pills; did he know about this helpful effect of his pills? Did he think of this possibility? Did he intentionally refrain from helping? Was a charitable act done mainly for altruistic reasons or mainly for other reasons like idleness, self-representation, desire for reward and acknowledgement? 4. In other cases we want to know the intention for judging the agent's responsibility and rational capacities on this basis: Is the agent's reasoning intelligent or confused? How far-reaching and clever is his planning of the action's details and their consequences? Was the goal intention suggested to and drubbed into the agent by inner voices or by an authority on which she is dependent? Was some desire (nearly) irresistibly strong (e.g. because of an emotion or corporal craving)? - The various answers to these questions make a difference in how to deal with these agents. This array of questions shows that we may be interested in more or less detailed knowledge about the agent's intention, which as a consequence may require more or less precise and costly means for interpreting the action.

A second, already derived, however still classical case, is that we do not even know the action but only some of its consequences and want to identify action itself, the agent, the underlying intention and/or further consequences. One key idea to resolving this problem is to take all (or some of) the known consequences as intended consequences of an action and then to try to reconstruct a fitting intention and action. Some variants of this epistemic situation are these: 1. We have a written text and want to understand its so far unclear meaning; or we want to know what has been said in the missing part of an incomplete text. 2. The police knows about the consequences of a crime - a person murdered or things missing, a forced door etc. - and wants to know who committed it and for what reason. 3. Archaeologists have found a strange iron hook or some other object and want to know what purpose it served and who made and used it.

A third, less frequent case is that we want to know the intention - in particular, of course, the implementation intention but also the goal intention or other pieces of the comprehensive intention - as a basis for predicting another person's (probable) behaviour. This may help us 1. to coordinate behaviour (e.g. to meet this person or to speak to her - if she wants e.g. to go to a certain restaurant), or 2. to prepare for the other person's action (if the landlord puts my house on the market I want to be the first person to make a bid; if the mother-in-law comes for an unannounced visit the daughter-in-law might want to have the house clean and tidy) or 3. to prevent the intended action, even to prevent a crime (the other person has planned to kidnap someone, to plant a bomb etc.). - But, how can we make predictions about actions with the help of an action interpretation, which already presupposes some knowledge about the action to be interpreted? The solution to this puzzle is that we are speaking of at least two actions. One action has already been accomplished, its underlying intention contains beliefs, general desires, overarching goal intentions and the like which are also parts or premisses for parts of the second comprehensive intention, still to be executed. In particular, sometimes we can make the desired predictions because many actions are intrapersonally coordinated; several smaller actions serve to reach one bigger aim; and the initial actions may make sense only if they are undertaken with a certain aim that, however, can only be realized with the help of further steps.

3. Several Conceptions of Action and Action Interpretation

The above list specifies the most important kinds of knowledge we want to acquire by means of action interpretations. Now the question is: which conception of action and of action interpretation can provide this desired knowledge? My two-part answer to this question is this. 1. Only an intentional causalist concept of actions, which conceives them as a behaviour caused and controlled by intentions, can be the basis for providing the desired knowledge. 2. And the most precise form of an action interpretation is a causal interpretation that tries to reconstruct the mental causes of the action, i.e. the deliberative case story or its essence, which is the intention.

Given the first part of the answer, the second part may be rather obvious. Some reasons for the first part of the answer are these. In the third of the above cases, i.e. when the action interpretation shall help to predict another person's behaviour, it is rather obvious that we need a causal conception of the action and a causal reconstruction of its underlying intentions. The reason for this is that in order to make the prediction we need to know the real causes of this behaviour; and we can know these causes because the central cause is a common cause of an action already executed and of the action to be expected. A mere *rational* interpretation of the preceding action would not allow us to make the prediction because the rationale does not necessarily correspond to mental facts and thus does not provide the information about such factual circumstances that can serve as premisses for a predictive inference. - In the other two cases the need for a causal interpretation is less obvious but nonetheless there. Retrospective, ascriptive responsibility is mainly a causalist concept. We ascribe responsibility to those features, states and events that are crucial and, for society, rather easily controllable jacking points for a possible social intervention to change the course of events. That a person is responsible for an event implies that a different action by this person would have prevented the event.[2] (Of course, this is only a necessary and not a sufficient condition for retrospective responsibility.) We establish responsibilities afterwards to learn how to prevent (or to repeat) similar events, or to punish (or reward) its authors with the aim of discouraging (or encouraging) other persons who might be in an analogous situation of responsibility. If the action interpretation aiming at establishing responsibilities were not identifying real mental causes of the behaviour the whole practice of changing future behaviour by intervening on its causes would not work. - Likewise, when we want to know the real intention behind an action or artifact - the intended meaning of a text, the last will of a person, the function of an ancient relic etc. - we are looking for an intention and sense that are *authentic*, i.e. can really be ascribed to the respective person who is responsible for them. We do not want to know if a certain meaning and sense would be *rational* but if it was *his* or *hers*. In our next argumentative step, identifying the authentic intention and the causing intention, of course, is not a logical necessity. However, up to now no other comprehensible and useful conception of a person's authentic intention has been developed. - Similar considerations hold if we want to judge the rationality of a person's action and intention. In such cases we do not want to know whether there may have been rational reasons for acting as the agent did but what reasons actually guided the agent and whether they were rational. Knowing this may also help to predict the agent's subsequent action.

The main rivals of the causalist conception are intentionalist conceptions of action and action interpretation. There are two main forms of intentionalism. The first one goes back to neo-Wittgensteinian ideas and says that interpreting behaviour as actions is just to describe it in a different language; instead of saying, for example, that 'Peter's hand is making such and such a movement' one says 'Peter is writing the word such and such'. According to neo-Wittgensteinianism, this language is based on necessary means-end relations, and by describing what happens it already ascribes intentions (cf. e.g. Stoutland 1976; 1989; Sehon 2005). The second form of intentionalism takes action interpretations to be assignments of objective (and subjective) reasons to an action: in the light of these reasons the action was rational. The precepts that are used for assigning such reasons are schemes of rational deliberation which are taken to be analogous to logical truths. The two most important schemes proposed for this purpose are practical inferences and rational decision theory (cf. e.g. Wright 1971, ch. III; 1972).

In the present context the most important objection to all these conceptions is that they do not provide the knowledge that we are after when conducting an action interpretation. The neo-

Wittgensteinian approach attributes an allegedly necessarily inherent goal intention to the agent, whereas the rationalist approach provides a rational reason, but neither of them tries to discover the intention for which the agent *really* performed the action, which implies that this intention caused the action. In addition, the presuppositions of the neo-Wittgensteinian approach are false. There are only very few cases with a necessary biconditional relation between goal intention and behaviour, i.e. where the behaviour is a necessary means for the goal intention and a certain goal intention the only possible intention behind that kind of behaviour. Furthermore, descriptions like 'Peter is writing the word such and such' are resultative, they describe an internally caused behaviour by way of its results, they are not necessarily, and therefore not analytically, intentional; the name "Peter" could refer e.g. to my cat walking over my computer keyboard thereby writing the respective word. The rational interpretation schemes on the other hand are far from being logical truths; and they are not general empirical truths either. Practical inferences rather represent only the most simple form of a deliberation; whereas the precepts of rational decision theory describe only the scheme of a rather highly developed and risk neutral deliberation; they do not represent simpler or more sophisticated or risk-avoiding or risk-seeking styles of decisions.

However, intentional causalism has been criticized too. Three prominent, and in our context relevant, objections are the following. First, an intentional causalist conception of action and intentionality has to resolve the problem of deviant realizations of intentions. According to this conception, the conditions for bringing about *p* intentionally include (i) that the agent intended to bring about *p* and (ii) that this intention caused *p*. However, there are cases where these two conditions are fulfilled but, nonetheless, the realization of *p* is not intentional because it has been brought about in a strange, deviant way; and this shows that the just mentioned intentional causalist conditions for intentionality are at least insufficient. Second, any causalist conception of action and intentionality presupposes a covering law that connects intention and behaviour. However, such a covering law, as the critics say, has not yet been found and cannot be found in principle. Third, recent experiments in neurophysiology seem to show that the whole traditional idea of intentions causing corresponding actions is false. For example, Libet concludes from his experiments that we can measure readiness potentials in the motor fields and on the vertex of the brain that clearly predict which action will be done (if an action is performed at all) before the respective intention is formed; therefore this intention cannot be the decisive cause of the action (Libet 1985; cf. also: Haggard & Eimer 1999). And Wegner reports many evidences of dissociation between our feeling of consciously acting and voluntarily controlled doing itself, i.e. cases where people feel that they are intending and performing an act that they in fact are not doing or, conversely, are not intending an act that they actually are doing (Wegner 2002). This is supposed to show that intentions are not the causes of actions but have a quite different role.

My answer to the first objection (deviant realizations of intentions) is that this objection is correct in stating that the fact that an intention causes the intended event is not sufficient for making the bringing about of this event intentional. However, this does not prove intentional causalism to be false; it shows only that, besides the conditions (i) and (ii), at least one further condition is missing. And this condition requires that the intention bring about the event in a controlled manner, first the action itself and then the consequences. For the piece between the action and consequences, control e.g., very roughly, requires (i) that the agent have some ideas about how the action leads to the intended consequence and (ii) that these ideas be more or less true.[3] So there is an answer to the first objection. An answer to the second objection will be provided below (sect. 5) by briefly describing such a covering law.

The third objection requires a more extensive response. The subjects' task in Libet's experiments was to quickly move a finger in an arbitrary moment whenever they felt an urge to do so; in addition they had to observe when these urges developed and report their exact timing measured by the position of a rotating light spot (Libet 1985, 530; 532). Libet then measured and compared the timing of the readiness potentials, the urges and the doing and found that they occurred in this order. Now, first, an urge to act is not an intention; a mother, bathing her baby, may feel an urge to drown the baby but never form the corresponding intention. Neither readiness potentials nor urges lead automatically to the corresponding doing; so it may well be that there was an intention, probably inspired by the urge but not simply resulting from it, that was the really decisive cause of action (Mele 2007). Second, Libet does not consider the main intention, namely the subject's general intention to follow the experimenter's instruction and to perform a long series of finger flexions etc. It is quite unlikely that this intention was a consequence of a readiness potential in the motor field of the right index finger. Third, the decision to flex one's finger now or one or two seconds later is so trivial and unimportant that the subjects perhaps formed only *one* intention, a general intention to flex a finger whenever feeling the urge to do so. Of course, such an intention, if (non-deviantly) effective, would be sufficient to turn the whole series of finger flexions into intentional actions. So, altogether, Libet's experiments by no means prove that intentions are not the decisive cause for actions.

Wegner, on the other hand, when trying to establish that "conscious will" is an illusion, uses this term in a quite peculiar sense, namely as denoting the experience that one's thoughts cause the thought of behaviour (Wegner 2002, 3; 14; 29 f.; 65 f.). What he mainly shows then is that these experiences are not a direct consequence of the corresponding causing but the results of our experiences and our cognitive considerations about the causal relations between them, which are similar to other cases where we try to establish causal connections. I think Wegner is right in this, and the material he presents is very interesting and convincing. It confirms, specifies and nicely illustrates a theory that some philosophers of action (including myself) had already held before, which, among other things, says that people have to learn (and extend) which kinds of behaviour are consciously controllable and even which kinds of mental states can cause such behaviour after representing it, i.e. which kind of mental states are intentions. Of course, all this is by no means in contrast to the really interesting hypothesis that the will or intention is a mental state that first represents the subject's behaviour and then causes it. However, Wegner goes on to claim that intentions (he calls them mainly "thoughts") are not the actual cause of action but only a by-product of an unconscious common cause of action (Wegner 2002, 67-69).[4] Unfortunately, Wegner does not substantiate this part of his theory that much. His main proofs are again Libet's readiness potentials and, secondarily, automatic actions (Wegner 2002, 49-59), and he does not specify what the unconscious common cause is - e.g. the readiness potentials? First, intentional automatic responses, indeed, are not preceded by singular intentions. But this is not necessary for making the responses intentional; the general intention, formed at the beginning of a series of automatic responses and functioning as the structuring cause though not as the triggering cause, is sufficient. And this general intention in fact is there and effective. Second, it is quite unlikely that the readiness potentials are such a common cause of the intention and of the action. Wegner himself says that we do not know what specific unconscious mental processes the readiness potentials in the motor area represent (Wegner 2002, 55). They are very closely connected to the single executive organs, and this makes them implausible candidates for having the decisive role in choosing action. This holds because such a decision has to take into consideration various options and their respective consequences, which probably are not represented in a single motor field. Already the frontal lobes seem to have much higher control functions. Third, our conscious decisions are often rather complex, taking into consideration i. various and ii. sometimes very complex options (consisting of many steps to be undertaken), iii. their consequences, iv. the subjective value of such consequences, and v. the comparative total value of the options - think of the decision to buy a house or shares, to write a certain book etc. In fitting the situation and providing good results for the subject our real actions seem to reflect such considerations; and it is difficult to explain this fittingness without resorting to the assumption that the real decider, either the conscious deliberation and decision or Wegner's "unconscious cause of action", did not use or have these considerations and information available and that the considerations were decisive. This still does not exclude that the unconscious cause of action also had such information etc. and really took the decision; but on the basis of what we know about brain functions and on evolutionary grounds it is not very likely that two parallel systems of decision-making using the same information have developed. So, without further evidences, it is much more likely that the conscious decision after the deliberation was the real decider rather than another unspecified "unconscious cause of action".

4. Simple Action Interpretations

As stated above, the aim of an action interpretation is to find out the intention or pieces of an intention underlying a behaviour. This is done on the basis of knowledge about the behaviour or the consequences of such behaviour; in addition, we may have further knowledge about the beliefs or other attitudes of the agent as well as knowledge about further circumstances. Given this fairly general characterization of action interpretations there can be several methods for interpreting actions. On the one hand, there are simple, fast and cheap action interpretations, which can be applied if we have already much information about the action, if the action is of a rather well-known type and if we do not want to know many details of the comprehensive intention. At the other extreme are complicated, costly and time-consuming action interpretations for cases where we know little about the action, have little background knowledge, where the action is of a rather unique type or where we want to know specific or many details about the comprehensive intention. Let us begin with the simple case.

In the most simple case we know the behaviour and the circumstances, and we know fairly general statistical regularities that connect this type of behaviour with a certain kind of intention. In such cases the action interpretation can take the form of a simple statistical inference:

P1: *s* eats - i.e., according to the resultative meaning explained above, *s* internally produces movements that result in ingesting food.

P2: People who eat, extremely frequently do so intentionally, which implies that they intended to eat.

P3: There are no clues to the contrary.

T: Therefore, with the utmost probability, *s* intended to eat.

Being statistical, such inferences are, of course, subject to the usual restricting conditions of probabilistic and non-monotonic reasoning, in particular that the reasoning subject must not have contrasting or more specific evidence. This limitation is indicated in premise P3. Inferring in this way is so easy and can happen so quickly that we do not have to execute it explicitly (of course internally); we immediately assume that such a behaviour is intentional. This holds e.g. if we observe activities like eating, drinking, dressing, washing, shopping, making love, cooking, driving, parking a car, writing etc. to which we can assign a result that, according to usual human attitudes, has some positive value. The general presupposition is (approximately): that

IH1: Intentional movement law: Internally caused movements of our limbs, head, mouth and body during vigilance (except ...) are nearly always intentional, whereas breathing, yawning, blinking (except ...) etc. are not.

The precise extensions of this presupposition differ for different epistemic subjects as well as for different agents. Sometimes they enclose false positives like holding the withdrawal of the hand from a hot object to be intentional, though this is mostly reflexive, or holding the movements of a patient in the vegetative state to be intentional. With increasing experience epistemic subjects refine IH1 (transforming it into IH1.1 etc.) on the basis of more precise forms of action interpretation, thus raising the relative frequency of IH1. The intentional movement law also covers behaviour that we - at least presently - do not really understand; we might e.g. observe a worker pressing buttons on a big machine unknown to us, which sometimes makes noise but produces no visible output; in such a case we would infer that the worker in any case is acting intentionally, that he is operating the machine, probably earning his livelihood by doing so etc. but we would not know the goal intention of his single actions.

In other cases, where we can identify a predictable action consequence that for many people has a positive value, we use the general presupposition together with this information in order to infer that this predictable consequence probably was an intended goal. So the additional general premise used here is this:

IH2: Goal intention law: If a person *s* acts intentionally and this action produces a rather easily predictable consequence *c* that for *s* has a positive value then, nearly always, *s* brings about *c* intentionally.
This goal intention law extends our possibilities of action interpretation considerably. In particular it can help to refine the intentional movement law (IH1). We may find out that a certain kind of movement is not intentional because we cannot attribute any goal to it - trembling, a tic etc.

The major part of our common action interpretations can be reconstructed as the use of the simple methods described so far. However, many limitations remain. The general hypothesis IH2 is only a statistical hypothesis with a high frequency; it says nothing about not so easily predictable consequences; and we have to know whether *s* cherishes the consequence. Because of these limitations, we need more sophisticated instruments for the interpretation of actions in more complicated cases. Such instruments will be presented in the following sections.

Before discussing these more sophisticated instruments a further standard case of action interpretation, namely the interpretation of speech acts, will be considered. Speech acts are special actions; and as actions they are caused by an intention. Therefore we can apply the general methods for action interpretation already described or still to be explicated here to speech acts as well. What is special about conventional speech acts, however, is that the shared meaning of a conventional language is constructed in such a way that the utterances, on the ground level, should express a precisely defined inner state of the speaker. Uttering an assertive with the proposition *p* should express the speaker's belief that *p*; uttering the respective interrogative should express a desire to know whether *p*; imperatives should express a desire (and the belief in the right to get this desire fulfilled) that someone will do something to make *p* true; etc. Of course, all this holds only on the ground level; on higher levels we have irony, role playing, fiction, implicatures etc., which, however, all presuppose the ground level. The minimal and again ground level consequence of such locutionary acts is that the hearer believes that the speaker believes that *p*, that she wants to know that *p* or wants *p* to be realized. Because these are easily predictable consequences of locutionary acts, which often have some positive value for the speaker, and because speaking (according to the intentional movement law) usually is intentional we can apply our general goal intention law (IH2) and attribute to the speaker the goal intention to make the hearer believe that the speaker believes that *p* etc. This easy way of attributing such intentions, subsequently, opens the possibility of a spiral of further higher order intentions: the speaker intends to make the hearer believe that the speaker wants the hearer to believe that the speaker believes *p* etc. (cf. Meggle 1981).

Language rules have been improved over the course of history in such a way as to reduce ambiguities; and speakers are educated to prevent ambiguities where they are possible. Nonetheless ambiguities exist, or (written or oral) texts may be mangled, faulty or distorted. The approach just touched nicely justifies the usual first order rule for dealing with such problems, i.e. the
ground rule of charity in semantic interpretation: In semantically interpreting texts, assume that an utterance has the semantical meaning which fits quite well to the known text and for which holds: in case of an assertive locutionary act, the speaker believes in the locution's propositional content;

in case of an imperative locutionary act, the speaker desires the locution's propositional content to be realized (and she believes that she has the right to get this desire fulfilled); in case of an interrogative locutionary act, the speaker desires to know whether the propositional content is true. This charity rule relies on the above mentioned basic rule of conventional language that an utterance, on the ground level, should express the speaker's appertaining inner state.

5. The Prerequisites of Complex Action Interpretations: Laws for Explaining Actions

Complex action interpretations imply more or less extensive causal explanations of the action by the underlying intention. These explanations are psychological statistical explanations. As such they presuppose a group of respective statistical laws. Although this is not the place to expound on such a decision psychology, we can, and must touch on it briefly.[5]

In empirical psychology by far the most accepted model of human choice is decision theoretical in the sense that the subject chooses between several options, he takes to be available to him, on the basis of assessing the values and probabilities of the options' implications, i.e. the options' advantages and disadvantages.[6] There are several main interpretations of the decision theoretical model. The most common interpretation is a black box model in the sense that it shall only capture the relation between the subject's desires and beliefs on the one hand and the resulting action on the other hand; it does not intend to say anything about the internal process mediating between them. However, exactly this black box approach has led to a gross impreciseness of the respective models; there are just too many ways in which subjects proceed from desires and beliefs to decisions (e.g. Harless & Camerer 1994). Furthermore, and particularly relevant in our context, hydraulic models do not say anything about goal intentions - there are just several positive consequences -; they do not distinguish between unintentional and intentional omissions; and they leave no room for false calculations. To avoid these failures, the decision theoretic model should include the internal steps and the internal result of the deliberation.[7]

A good way to connect the subject's "premisses", the deliberation process and the final result of the deliberation is to conceive this result as an optimality judgement that a certain option, from the personal point of view, is the best among the considered options and to conceive the deliberation as a reflective process with the aim of finding a true optimality judgement. According to this approach, the first, statistical but nonetheless rather strong empirical covering law, roughly, says:

IH3: Action execution law: 1. If a person *s* at t_- (i.e. before *t*) has formed an optimality judgement that to do *A* at time *t* is, from her personal viewpoint, the best of the considered feasible options,

2. if *s* has not retracted this belief between t_- and *t*,

3. at *t*, *s* is aware of the fact that it is now *t*, and

4. at *t*, *s* is capable of doing *A*,

5. then, in the vast majority of cases, *s* begins to do *A* at *t*.

The next step of the explanation regards the phase before the forming of the implementation intention, i.e. the phase of putting the various considerations about advantages and disadvantages for the various options together into forming the optimality judgement. In particular, a covering law for this phase should describe the subjects' dealing with probabilistic consequences and risk. Risk behaviour has been the object of a huge number of studies in economics and psychology. The results, in a certain respect, are disappointing. There is a great variety of procedures and criteria for aggregating the assumed advantages and disadvantages of the various options to the final optimality judgement. The ways of aggregation vary interpersonally and intrapersonally from one situation to another (overview: Payne, Bettman & Johnson 1993). They differ with respect to how many options and consequences are considered, which type of consequences is considered at all and in which order, how the various options are compared, how probabilities are weighted, etc. The aggregation methods, at least in part, seem to be inventions of the respective agent herself and subject to an assessment regarding their preciseness and costs. This means that, in a certain sense agents decide how to decide. Up to now, we can only recognize a tendency to optimize the decision procedure, i.e. to intuitively use such aggregation methods for which, in this particular situation, the balance between preciseness and decision costs could be optimum. However, at the moment the studies have not led to hypotheses that would would make it possible to predict the precise aggregation method used in a certain type of situation. Therefore, at this stage, our last resort with respect to laws about the aggregation method is to return to assume, as an acceptable approximation, that agents aggregate according to expected value. However, even this approximation does not say anything about which of the probably accessible information about the options were taken into consideration. We can only surmise that with the increasing subjective importance of the decision more and less obvious aspects of the options were taken into account.[8]

The third explanation step regards a systematically still preceding phase, namely the assessment of the action outcomes according to one's personal criteria (cf. Lumer 1997). On the one hand, we have more or less inborn or original criteria for intrinsically assessing such outcomes. On the other hand, during deliberation agents often use stored valuations of these outcomes, which in a more or less complicated genesis result from the application of the original criteria for intrinsic assessments. The use of stored valuations is due to the fact that the consequential chains from a possible action to its intrinsically relevant consequences often are very manifold and long; therefore, during our personal history, we store assessments of intermediate consequences with more and more far-reaching implications. The most important original criteria for the intrinsic evaluation of action consequences are, first, hedonistic and, second, feeling-induced intrinsic valuations. According to *simple hedonism*, the personal intrinsic value of feelings corresponds to the product of their (positively or negatively) sensed intensity and their duration. This hedonic criterion corresponds to Bentham's quantitative hedonism. Via stored assessments of intermediate action consequences, hedonic valuations are at the basis of most of our assessments of consequences; they generate rather stable valuations appropriate for long-term decisions. *Feeling-induced intrinsic desires*, on the other hand, emerge from present feelings and emotions and are a response to these feelings. Rage, for example, induces an intrinsic desire to punish the object with which one is furious; pity induces the intrinsic desire to have the other person's situation improved. Such intrinsic desires vanish as the feeling fades away. Feeling-induced desires are the basis of emotional acts; sometimes we regard such emotional acts as irrational precisely because the desires for which they are executed are so unstable.

The foregoing, of course, can only provide a rough idea of what kind of psychological laws can be used in action explanations. However, even after a specification and elaboration of these laws, we still need information about the agent's single beliefs and stored desires in order to explain actions. If we cannot ask the agent about them we can try to arduously extrapolate them from known facts about his life history. There is, however, still another important source of such knowledge, namely our knowledge about common knowledge and valuations in the agent's society or community.

6. Complex Action Interpretations - Inferences to the Best Explanation

Above I mentioned some limitations of simple action interpretations. They may presuppose a standard form of action; they may require very specific knowledge that is not available; the text of a locutionary act may be ambiguous, incomplete, faulty, distorted etc. If we want to have an action interpretation despite such obstacles, as the last resort we can turn to complex action interpretations on the basis of an inference to the best explanation.

The best known examples for this kind of action interpretation are "whodunnits": The detective has available several pieces of circumstantial evidence, including the immediate consequences of the criminal act, e.g. the corpse, and of the personal profiles of two or more suspects, their beliefs, possible motives, personal histories, finger prints etc. The strategy by which the detective determines the identity of the perpetrator is to construct coherent stories that contain all the circumstantial evidences and that explain the criminal act and its immediate consequences. For designing these stories the detective uses all the relevant information, which however usually is too riddled with holes to make a coherent story; so he fills the holes with more or less likely hypotheses, for which however, ex hypothesis, he does not have sufficient evidences. Every coherent story of this type may be called a "*(possible) construal*". If a commenced story turns out to lead to incoherencies, e.g. first an agent is supposed to have a certain belief that *p*, later on however she is supposed not to believe in *p*, then it is an *impossible construal*.

The first task in action interpretation by best explanation is to discard impossible construals, which first may have appeared to be possible.

The second task is to find possible construals. Because possible construals often contain merely hypothetical parts, in many cases there will be several or even many possible construals. Principally, the detective (or the interpreter) should find all possible construals. However, in many cases the number of possible construals is so high that the interpretation has to be restricted to the most probable construals. Unfortunately, sometimes even then the number of possible construals remains unmanageably high, which means that the information basis is so insecure and limited that the action interpretation does not provide positive results.

The third and final task is to find the best construal and to calculate its probability. From an epistemic viewpoint, the best construal is the most probable construal. However, at this point we have to distinguish between *a priori probabilities*, which are the probabilities we can assign to the hypotheses and construals before we begin with the calculation, and the *a posteriori probabilities*, which result from our calculation. First, we have to calculate the *a priori* probabilities of the single construals. The *a priori* probability of a construal is identical to the *a priori* probability that all the hypotheses occurring in the construal together are true. If all these hypotheses are mutually independent the *a priori* probability of the construal is identical to the product of the *a priori* probabilities of all hypotheses occurring in it. Let us consider a simple example: Let e_1, ..., e_n be the known relevant evidences with e_n being the immediate consequence of the criminal act (e.g. Black's corpse lying on the floor). Let h_{11} and h_{12} be two hypotheses that are needed to complement these evidences for getting a construal that with the help of some laws l_1 to l_m explains the final result e_n (e.g. h_{12} may say: Smith enters the room and shoots Jones).

construal 1	construal 2
$e_1, ..., e_{n\text{-}1}$	$e_1, ..., e_{n\text{-}1}$
h_{11}, h_{12}	h_{21}, h_{22}
$\underline{l_1, ..., l_m}$	$\underline{l_1, ..., l_o}$
e_n	e_n
$P_1(h_{11}) = 0.1; P_1(h_{12}) = 0.2$	$P_1(h_{21}) = 0.2; P_1(h_{22}) = 0.4$
$P_1(\mathrm{cl}_1) = P_1(h_{11}) \cdot P_1(h_{12}) = 0.02$	$P_1(\mathrm{cl}_2) = P_1(h_{21}) \cdot P_1(h_{22}) = 0.08$

Furthermore, let h_{11} and h_{12} be mutually independent and have the *a priori* probabilities $P_1(h_{11})$ = 0.1 and $P_1(h_{12})$ = 0.2; then the *a priori* probability of construal 1 is 0.1 times 0.2, i.e. 0.02 ($P_1(\mathrm{cl}_1)$=0.01). In the whodunit, one of the hypothesis, let's say h_{12}, should be an action description that the agent *s* committed the crime that led to e_n; and the other hypothesis, h_{11}, should be a description of the agent's intention. Similar considerations may hold for a second construal with hypotheses h_{21} and h_{22} and perhaps using some different laws; it may lead to the *a priori* probability $P_1(\mathrm{cl}_2)$=0.08. Now, if the *a priori* probabilities of all the construals have been determined, we can calculate the *a posteriori* probability of these construals. This calculation follows Bayes' Rule.

(1) $$P_2(h_\mathrm{i}) = \frac{P_1(h_i/e)}{{}_{j=1}\Sigma^m P_1(h_\mathrm{j}/e)}$$

(Transformed from: Eells 1982, 13) If all the hypotheses are independent of the known evidences, so that $P_1(h_j/e)=P_1(h_j)$ holds, this formula reduces to:

(2) $$P_{n+1}(h_i) = \frac{P_n(h_\mathrm{i})}{{}_{j=1}\Sigma^m P_n(h_j)}$$

The idea behind Bayes' Rule is this. At the beginning we have only the *a priori* probabilities of the several construals. Now, however, we add to this the information that exactly one of the construals must be true because these are all possible construals. With this information the probability space becomes considerably restricted. Where initially we had many possibilities with low probabilities, we now know that the probabilities of all the possible construals have to add up to 1; they make up the new probability space. Now, the essence of Bayes' Rule is that this probability of 1 is distributed among the remaining possibilities, i.e. the possible construals, in proportion to their *a priori* probabilities. In our example this means: because we had only two possible construals with the *a priori* probabilities of 0.02 and 0.08, the sum 1 of the *a posteriori* probabilities has to be assigned to the construals in the relation 0.02 to 0.08, which is 0.2 to 0.8. So we get the *a posteriori* probabilities $P_2(\mathrm{cl}_1)=0.2$, $P_2(\mathrm{cl}_2)=0.8$.[9] Finally, we can assign these *a posteriori* probabilities of the construals to all the hypotheses contained in them. So we get: $P_2(h_{12})=0.2$ and $P_2(h_{22})=0.8$. There is, however, one exception. One and the same hypothesis may occur in several construals. In such a case the *a posteriori* probability of this hypothesis is equal to the sum of the *a posteriori* probabilities of all the construals of which it is a part.

The interpretation procedure just explained by using the example of whodunits can also be used to interpret texts, because these are the results of speech acts - I have reconstructed an extensive text interpretation on this basis (Lumer 1992) -, or all kinds of actions proper. Therefore, these extensive action interpretations can be used in history, in psychology, in sociology and in other sciences that use action explanations.

On the other hand, the general method of interpreting known facts by means of inferences to the best explanation with the aim of knowing the causes of these facts is used also in many natural sciences: in geology, in biology, in astronomy, in chemistry, sometimes in physics etc. The characteristic of *action* interpretations as compared to these sciences is only that the former use as explaining laws, among others, the above sketched laws about the formation of intentions and the causing of actions. This continuity of complex action interpretations with other inferences to the best explanation also explains why there is no problem in taking, as is done in whodunits, the (rather immediate) consequences of an action as the central explanandum instead of the action itself. There is a continuity of causality.

7. Interpreting Moral Actions

Can the methods of action interpretation just explained be applied to interpreting moral actions? According to the methodology explained so far, this question can be specified and split into two questions: 1. Are moral actions subject to causal laws (not necessarily strict laws) in the Humean sense, or are they e.g. autonomous in a Kantian sense, i.e. subject to laws of Reason? 2. If moral actions are subject to a Humean causality, do the psychological laws of action and decision sketched above apply to them, or do different laws apply to them? I think the reply to both questions is in favour of a psychological normalcy of moral action; moral actions like all the other actions follow the usual laws of action and decision.[10] The only differences as compared to amoral actions are their special moral content and their motives.

Let us consider the first question first. The most important counter model to a causal explanation of moral actions is Kant's model of acting from pure Reason. According to this model, an *a priori* reflection of pure Reason, which establishes what we shall do, can also determine our will; this faculty to determine our will is called "pure practical Reason". The exact mechanism by which this happens, according to Kant's explanation in the "Critique of Practical Reason" (Kant 1788, A128-135), is that the moral law detected by pure Reason, in an *a priori* understandable way, first humiliates the self, because this law ignores the subject's personal inclinations, however in a second step generates a kind of admiration for this law, i.e. the respect for the moral law, which then is a motive for obeying this law. The general idea of this model is that pure Reason's *a priori* determination of what has to be done in an also *a priori* understandable way generates the respective motivation to act accordingly. However, a first objection to this model, the so called "content skepticism about practical reason" (Korsgaard 1986, 311), says that pure Reason cannot determine what we should do; pure Reason can establish that an action *a* has certain *a priori* properties; but it cannot establish which of these properties are practically relevant in such a way that having this property is a reason to perform this action; in brief: pure Reason cannot establish practical relevances.[11] Still more important in our context is a second objection, the so called "motivational skepticism about practical reason" (Korsgaard 1986, 311). There are several versions of this objection; the clearest and strongest I think is this. Suppose pure Reason has established that the property *F* is practically relevant and decisive in the sense that if we find that an action *a* has this property *F* we should do *a*. Kant assumes that these two insights can determine our will, i.e. our executive intention, in the sense of bringing about the respective motivation. Now, the only clear meaning of "to determine the will" or of "bringing about something" is that of a causal relation in the Humean sense. This, however, presupposes empirical regularities that a cognition of the type that a certain action is *F* leads to a motivation to do this action. And such a regularity would rely on our mental mechanisms - and not on a decision of pure Reason. In particular, if we were to react to the cognition of the moral law with humiliation and respect this would be an empirical and causal reaction; of course, from a logical point of view, we could also ignore this cognition or react with anger about this law and strongly disdain and refute the law.

This failure of an *a priori* approach to moral motivation and action, however, does not exclude that there are *empirical* mechanisms, fairly neutral with respect to the content of our moral insights, that provide motivation to act morally after having found out that a certain action has the morally relevant property *F*. To repeat, this would be an empirical mechanism. This brings us to our second question: what are the causal mechanisms of moral action.[12] Indeed, there is such a mechanism providing motivation to moral judgements; traditionally it has been called "*conscience*". For immature persons conscience may simply be a fear of socially induced punishment as a consequence of acting immorally or, to the converse, hope for recognition. In mature persons this external conscience is substituted (or at least complemented) by an *internal conscience*, and internal sanctions and recognitions. The internal sanctions and recognitions consist in a positive or negative alteration of our self-esteem as a consequence of a change in our moral self-appraisal. Low or negative self-esteem is an unpleasant feeling and high or positive self-esteem is a pleasant feeling. Therefore, the prevision of such feelings as a consequence of our acting immorally or morally causes a normal hedonic motivation to improve - or at least not to worsen - one's feelings by acting morally. So this mechanism falls under what has been said above (sect. 5) about intrinsic desires. This mechanism of an internalized conscience is fairly neutral with respect to the content of moral motivation; i.e. a low or high moral self-assessment leads to the respective feelings and motivation - irrespective of the criteria underlying these self-assessments. This, however, does not imply that the subjective acceptance of such criteria is the consequence of *a priori* reasoning and not again subject to empirical mechanisms. Here, however, we have to leave this question open.[13]

Acting on grounds of conscience is acting morally in the strict Kantian sense, that the motivation follows the moral judgement, which is the origin of the respective action. Of course there are many motives and desires to act *in accordance* with morals, which do not lead to acting morally in this strict sense. There are rather *accidental motives*, like the desire to earn one's living by being a social worker, the spirit of adventure of a development worker, the desire to feel one's power in organizing international aid etc. There are *motives more bound to moral contents* like desiring to cooperate for reasons of reciprocation, wanting to avoid social sanctions etc. As can easily be recognized, these two groups of motives fall under the explanation scheme sketched above insofar as these desires can be traced back to intrinsic hedonic desires. A third group for acting in accordance with morals consists of *self-transcendent motives*, which often have a morally desirable content, namely love, creative expansion of the self by creating socially cherished works and identification with a larger collective. Very important motives behind many of these self-transcendent motives are pride in one's works or in one's community and the experience of power - which again can be reduced to intrinsic hedonic desires to have these kinds of pleasant feelings. The motivational part of love, on the other hand, is not *one* motive but a conglomerate of several motives, in particular a specialized or accentuated sympathy and in case of one's children also of a creative expansion. A last and, for the justification of morals, very interesting group of motives for acting in accordance with morals are *motives near to morals*. The most important motives near to morals are sympathy and respect for other beings (which must not be confused with the Kantian respect for the moral law). Sympathy and respect, first and foremost, are emotions, namely feeling some sort of pain or joy as a consequence of believing another person to be badly and well off, respectively, or feeling some sort of admiration or sense of cherishing and caring for the object of one's respect. There are two ways by which these emotions can lead to motivation. The first one is again hedonic. Positive sympathy and respect are pleasant feelings, whereas pity is an unpleasant feeling. Therefore, we can improve our hedonic state by altering the situation of other beings in such a way that we have the positive and avoid the negative emotions near to morals. The second way originates in the fact that sympathy and respect are emotions and as such they can induce *intrinsic* desires. Strong sympathy can induce the intrinsic desire to improve or, in the case of positive sympathy, to protect the other's well-being. Respect can induce the intrinsic desire to protect and conserve the respected being and to leave room to its proper development. So the second way falls under what has been described above as "feeling-induced intrinsic desires".

This rush through the various motives for acting in accordance with morals reveals that acting morally does not constitute a separate form of acting. Of course, there are special motives for acting morally, however these are captured by the general framework described above (sect. 5). Therefore, we can also apply the above explained method of action interpretation (sects. 4 and 6) to moral actions. Doing this, we may find out which of the motives listed above had which weight in the decision - usually there is a melange of motives behind such decisions - and, thereby, find out whether moral motives, motives close to morals or, on the other hand, crude self-interest were dominant in a given decision.

Generalizing these results still further, it seems that the above outlined psychological laws of human intention formation as well as the simple and, still more, the complex and detailed method for interpreting actions can be applied to all forms of human action.

References

Bratman, Michael E. (1987): Intention, Plans, and Practical Reason. Cambridge, Mass.; London: Harvard U.P.

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.

Eells, Ellery (1982): Rational decision and causality. Cambridge: Cambridge U.P.

Frankfurt, Harry G. (1969): Alternate Possibilities and Moral Responsibility. In: Journal of Philosophy 66 (1969). Pp. 829-839.

Haggard, Patrick; Martin Eimer (1999): On the relation between brain potentials and the awareness of voluntary movements. In: Exp. Brain Research 126. Pp. 128-133.

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

Kant, Immanuel (1788): Critique of practical reason. (Kritik der praktischen Vernunft.) In: Idem: Critique of practical reason and other writings in moral philosophy. Translated and edited with an introduction by Lewis White Beck. New York: Garland ^{1}1949; ^{2}1976.

Koehler, Derek J.; Nigel Harvey (Hg.) (2004): Blackwell Handbook of Judgment and Decision Making. Oxford: Blackwell

Korsgaard, Christine M. (1986): Skepticism about Practical Reason. Reprinted in: Eadem: Creating the Kingdom of Ends. Cambridge: Cambridge U.P. 1996. Pp. 311-334.

Libet, Benjamin (1985): Unconscious cerebral initiative and the role of conscious will in voluntary action. In: Behavioral and Brain Science 8. Pp. 529-566.

Lumer, Christoph (1992): Handlungstheoretisch erklärende Interpretationen als Mittel der semantischen Bedeutungsanalyse. In: Lutz Danneberg; Friedrich Vollhardt (eds.): Vom Umgang mit Literatur und Literaturgeschichte. Positionen und Perspektiven nach der "Theoriedebatte". Stuttgart: Metzler. Pp. 75-113.

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

—— (1999): Handlung / Handlungstheorie. In: Hans Jörg Sandkühler (ed.): Enzyklopädie Philosophie. Vol. 1. Hamburg: Meiner. Pp. 534-547.

—— (2000): Rationaler Altruismus. Eine prudentielle Theorie der Rationalität und des Altruismus. Osnabrück: Universitätsverlag Rasch.

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

—— (2002/03: Kantischer Externalismus und Motive zu moralischem Handeln. In: Conceptus 35. Pp. 263-286.

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

—— (2007): An Empirical Theory of Practical Reasons and its Use for Practical Philosophy. In: Christoph Lumer; Sandro Nannini (eds.): Intentionality, Deliberation and Autonomy. The Action-Theoretic Basis of Practical Philosophy. Aldershot: Ashgate. Pp. 157-186.

—— (2008): Abwegige Absichtsrealisierung und Handlungssteuerung. Eine intentional-kausalistische Erklärung. In: Internationale Zeitschrift für Philosophie. Pp. 9-37.

Meggle, Georg (1981): Grundbegriffe der Kommunikation. Berlin; New York: de Gruyter.

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

—— (2007): Free Will. Action Theory Meets Neuroscience. In: Christoph Lumer; Sandro Nannini (eds.): Intentionality, Deliberation and Autonomy. The Action-Theoretic Basis of Practical Philosophy. Aldershot: Ashgate.

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

Sehon, Scott Robert (2005): Teleological Realism. Mind, Agency, and Explanation. Cambridge, Mass.: MIT Press.

Stoutland, Frederick (1976): The Causal Theory of Action. In: Juha Manninen; Raimo Tuomela (eds.): Essays on Explanation and Understanding. Dordrecht; Boston: Reidel. Pp. 271-304.

—— (1989): Three Conceptions of Action. In: Herbert Stachowiak (ed.): Pragmatik. Handbuch pragmatischen Denkens. Vol. III: Allgemeine philosophische Pragmatik. Hamburg: Meiner. Pp. 61-85.

Wegner, Daniel M. (2002): The Illusion of Conscious Will. Cambridge, Mass.; London: MIT Press.

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

—— (1972): On So-Called Practical Inference. In: Acta Sociologica 15. Pp. 39-53.

[1] For the following cf.: Lumer 1999, sects. 6-7.

[2] This is a special version of the principle of alternate possibilities (PAP), which says that someone is responsible for what he has done only if he could have done otherwise. Frankfurt has criticized PAP (Frankfurt 1969), and his critique has initiated a huge debate, which we cannot dwell on here. However, a proposal that grants the prima facie plausibility of PAP as well as that of Frankfurt's counter-example is this. PAP normally holds because it shall guarantee the personal accountability for events; however it does not hold in exceptional cases, especially not in cases of overdetermination, for which Frankfurt's Black-Jones story is an example. In such cases the overdetermined event relies on the decisions of several agents, who are all responsible for it though not one of them could have prevented it individually.

[3] This is not the place for further explanations of the required kind of control; I have done this elsewhere: Lumer 2008.

[4] Wegner does not mean this claim in the sense of epiphenomenalism, i.e. holding that not the intention but only its physical supervenience base causes the behaviour - in fact he does not reflect such philosophical declinations of mental causation -, but in the stronger sense that even the supervenience base of the intention

does not cause the action. So someone adhering to epiphenomenalism would have to characterize Wegner's model as 'double epiphenomenalism': the conscious intention is only a mental epiphenomenon of a physical epiphenomenon of the real cause of action.

[5] More details can be found in: Lumer 2000 128-240; 428-521; Lumer 2005; Lumer 2007; Lumer 1997.

[6] Cf. Koehler & Harvey 2004, sect. I; III; Camerer 1995. The decision theoretical model already excludes some other empirical models of action proposed by philosophers or economists. *Theories of practical inferences* (defended e.g. by Georg Henrik von Wright) are too simplistic as they do not consider the possibilities of various options, secondary consequences or their values and probabilistic consequences. Theories of *decisions according to the strongest desire* (proposed e.g. by Robert Audi) do not consider the possibility that the strongest desire for an action may be outweighed by an aversions to this action or by the sum of several weaker desires for another action. *Theories of satisficing* (fostered e.g. by Herbert Simon or Michael Michael Slote), according to which we do not maximize desire fulfilment but only try to surpass a certain minimum level of satisfaction, have difficulties in explaining the etablishment of a minimum level and do not see the possibility that establishing a minimum level usually is a maximizing measure by which we to prevent the costs of a further, inefficient deliberation. *Cognitive judgement hypotheses* (advocated e.g. by Thomas Nagel, John McDowell, David McNaughton, Mark Platts), according to which mere cognitive judgements can motivate for action, until today have not been developed to the level of a theory, in particular they lack any general psychological law that could be testable, and they cannot explain why one particular truth out of the infinitely many truths about an action shall be practically decisive in the sense that recognizing it will motivate to action.

[7] Bratman, Mele and several other philosophers see intentions as mental attitudes *sui generis*, irreducible to other kinds of mental attitudes like desires or beliefs (Bratman 1987; Mele 1992). One problem of this approach is that here again the final result, i.e. the intention, is detached from the deliberation process, there are no "laws of reasoning" leading from the considerations undertaken during deliberation to the intention. And this leaves a gap in the action explanation.

[8] For a more precise but much more complex approach see: Lumer 2005.

[9] $P_2(\mathrm{cl}_1) = P_1(\mathrm{cl}_1)/(P_1(\mathrm{cl}_1)+P_1(\mathrm{cl}_2) = 0.02/(0.02+0.08) = 0.2$;
$P_2(\mathrm{cl}_2) = P_1(\mathrm{cl}_2)/(P_1(\mathrm{cl}_1)+P_1(\mathrm{cl}_2) = 0.08/(0.02+0.08) = 0.8$.

[10] For extensive justification of this claim see: Lumer 2002.

[11] This and the following criticism are set out in more detail in: Lumer 2002/03.

[12] The following list of moral motives and of motives to act in accordance with morals is taken from and explained in detail in: Lumer 2002.

[13] The ontogenetic development of moral criteria is outlined in: Lumer 2002, sect. 7.