Philosophical Studies (2006) 128:201–227
DOI 10.1007/s11098-005-4063-x

HEINRICH WANSING

DOXASTIC DECISIONS, EPISTEMIC JUSTIFICATION, AND THE LOGIC OF AGENCY

ABSTRACT. A prominent issue in mainstream epistemology is the controversy about doxastic obligations and doxastic voluntarism. In the present paper it is argued that this discussion can benefit from forging links with formal epistemology, namely the combined modal logic of belief, agency, and obligation. A *stit*-theory-based semantics for deontic doxastic logic is suggested, and it is claimed that this is helpful and illuminating in dealing with the mentioned intricate and important problems from mainstream epistemology. Moreover, it is argued that this linking is of mutual benefit. The discussion of doxastic voluntarism directs the attention of doxastic logicians to the notion of belief formation and thus to dynamic aspects of beliefs that have hitherto been neglected. The development of a formal language and semantics for ascriptions of belief formation may contribute to clarifying the contents and the implications of voluntaristic claims. A simple observation concerning other-agent nestings of *stit*-operators, for instance, may help illuminating the notions of making belief and responsibility for beliefs of others. In this way, *stit*-theory may serve as a bridge between mainstream and formal epistemology.

1. INTRODUCTION

The justification of scientific as well as everyday belief (knowledge), usually referred to as doxastic (epistemic) justification, is a central and controversial issue in mainstream epistemology; for a recent survey see (Lammenranta, 2004). In particular, the question whether doxastic justification might be analyzed in deontological terms is contentious. M. Steup (2001, p. 135) characterizes the deontological conception of epistemic justification (alias deontologism) as follows:

> α is justified in believing that p at t if and only if it is epistemically responsible of (permissible for) α to believe that p at t (*notation slightly adjusted*).

Some epistemologists hold that an internalistic understanding of epistemic justification receives considerable support from deontologism (Plantinga, 1993; Goldman, 1999), so that internalism may be deprived of much of its support by arguments against deontologism. Perhaps the most familiar line of attack against deontologism consists in claiming that the deontological conception presupposes doxastic voluntarism (also called volitionalism) and that doxastic voluntarism is false, see (Alston, 1988). According to Shah (2002, p. 436), "[i]t is common for philosophers to claim that doxastic voluntarism, the view that an agent can form beliefs voluntarily, is false." This is, however, a quite distorted description of the views philosophers have expressed concerning doxastic voluntarism. It ignores a relevant part not only of the recent literature (see, for example, (Steup, 2000), (Ginet, 2001), (Ryan, 2003)) and declares a consensus where no agreement exists.[1]

In this paper, I will not discuss what is maybe the most fundamental question concerning doxastic voluntarism, namely: What exactly does the doxastic voluntarist claim? Doxastic voluntarism has been characterized by philosophers in many ways, not all being equivalent and some being quite nonspecific or even unclear. One popular way of characterizing doxastic voluntarism is in terms of control. R. Audi (2001, p. 93), for example, presents doxastic voluntarism as "the thesis that belief is sometimes under direct voluntary control." In terms of decisions to believe, the corresponding thesis would be that a doxastic subject may sometimes decide to believe something, in the sense that the decisions directly result in beliefs. In terms of belief formation (understood as an action of belief acquisition), the thesis would be that a doxastic subject may sometimes form a belief (as a result of deciding to form it).[2] In the present paper, I shall leave the characterization of doxastic voluntarism at this level of explication.

It is well-known that in deontic logic a distinction is drawn between theories of *tun sollen* (*ought to do*), according to which deontic notions apply to actions, and theories of *sein sollen* (*ought to be*), according to which deontic notions apply

to states of affairs or static situations. If deontologism is attacked by claiming that it presupposes doxastic voluntarism, which is deemed to be wrong, then a theory of *tun sollen* is presupposed. If it is states of affairs that are permissible, forbidden or obligatory, a critique of doxastic voluntarism just misses the point. It seems, however, and I agree on this with Audi (2001) and Engel (1999), that, a number of dynamic locutions about believing notwithstanding, believing itself is not an action type and that "to believe" is not an action verb. This means that the thesis of deontologism ought to be re-stated:

α is justified in believing that p if and only if α is permitted to form (or voluntarily acquire) the belief that p.[3]

It is not the aim of this essay to contribute to discussing whether deontologism in fact presupposes doxastic voluntarism or to present a detailed analysis or defense of doxastic voluntarism. I shall, however, critically consider three arguments that are meant to support doxastic anti-voluntarism. In each case, a certain problem in the debate about doxastic voluntarism is identified and discussed. The results of these considerations are then taken into account in a formal theory of belief formation. In the light of the discussion of the anti-voluntaristic arguments, I shall introduce a formal language and its model-theoretic semantics. This language extends the language of the modal logic of agency called *dstit*-theory (where "*dstit*" abbreviates "deliberately sees to it that"). This theory has been developed by Nuel Belnap, Michael Perloff, Ming Xu, Franz von Kutschera, and John Horty, see (Belnap et al., 2001) and references therein. Into the language of *dstit*-theory standard modal belief operators are introduced, and in order to discuss normative concepts, the language may be further extended by deontological modalities, for example along the lines of (Wansing, 1998) or (Horty 2001). The usefulness of the formal language and its semantics is then exemplified by discussing the notion of deciding to know, the notion of making believe and the formulation of deontologism.

I take it that both the anti-voluntarist and the voluntarist should appreciate a formal theory that contributes to making available a precise and unequivocal statement of certain problems surrounding doxastic voluntarism and thereby might help clarifying the problems themselves.

2. THREE ARGUMENTS AGAINST DOXASTIC VOLUNTARISM

2.1. *The Argument from Intention*

A familiar argument against doxastic voluntarism is based on the view that genuine acting, and decision making in general, is intentional. A variant of this argument from intention is presented by Feldman (2001, p. 85), who claims that "on the compatibilist view, to be a voluntary action, an action must be caused by an appropriate intention to perform that action." Feldman does not explain how he conceives of intentions and how exactly the formation of an intention is supposed to be involved in decision making. However, for him it seems to be clear that "we typically don't form intentions to form beliefs and form them as a result." The restriction to typical cases is important here, because Feldman concedes that there are odd cases in which agents do intend to form a belief and in fact form the belief as a result. He uses the argument from intention to argue that in cases in which belief is the result of weighing evidence, belief formation is involuntary. His conclusion, however, is quite general: "Believing may be the consequence of a deliberative process, but it is not voluntary behavior" (Feldman, 2001, p. 86).

Let α be an agent capable of decision making. According to Feldman, the central usage of the notion of a decision can be found in sentences of the form:

(1) α decided to Q.[4]

It is claimed that (1) logically implies:

(2) α formed an intention to Q.

If, in cases of deliberating about what to believe, belief formation is a kind of action,

(3) α decided to form the belief that p

in Feldman's view therefore implies:

(4) α formed an intention to form the belief that p.

Feldman is certainly right that α's weighing the evidence for and against p is normally not accompanied by an intention to form the belief that p (or an intention to disbelieve that p, or an intention to neither believe nor disbelieve that p). He suggests that agent α's belief formation as the result of deliberation about the evidence for and against p is best described by:

(5) α came to the conclusion that p.

Since sentence (5) fails to be an action report, it seems that "[w]e don't decide what to believe on the basis of that evidence" (Feldman, 2001, p. 85). Thus, for Feldman, coming to a conclusion is not a form of deciding to form a belief.

The key-step in the argument from intention is the claim that (1) logically implies (2). Of course, one may hold that this implication is valid by definition, but I would like to argue that there are *unintentional decisions*. Pojman (1985, p. 39) requires for deciding to believe that "the acts of will which produce belief are decisions of which we are fully aware." What I intend to show is that an agent may make an unintentional decision while being aware of making a decision.

Before I come to the argument, let me make a remark on agentive sentences. According to the Stit Paraphrase Thesis, a statement q (in present tense) reports an action by agent α iff (up to an approximation) q is logically equivalent with

(6) α sees to it that q

see (Belnap et al., 2001, p. 7). If q is a retrospective report of an action by α, using past tense, we obtain

(7) α saw to it that q.

My consideration does not depend on the truth of the Stit Paraphrase Thesis. However, I will assume that if the

sentence "α formed the belief that p" is an action report, it is both adequate and useful to re-formulate it as

(8) α saw to it that α formed the belief that p.

My argument rests upon the assumption that genuine actions involve decisions. By a decision I understand an agent's choice between at least two options for acting. A choice is something an agent makes by acting.[5] Therefore, on the above assumption, (8) implies

(9) α decided to see to it that α forms the belief that p.[6]

It seems to be clear that there are *unintentional actions*. It may happen that an agent α sees to it that q without intending to see to it that q. (Note that I am not alluding to unintentional effects of intended actions but to unintentional actions.) Agents are performing unintentional actions all the time. In walking on the beach I may be aware of deciding where to put my feet. It would be mistaken, however, to assume that in walking on the beach I necessarily first form an intention to put my feet somewhere or even on a particular region of sand and then make a step. Unintentional, unplanned actions may involve conscious and may also involve unconscious decisions. I may, conscious of making a decision or not, decide to take the stairs instead of the lift, without first forming the intention to take the stairs.

Thus, let us suppose that in addition to (7) also the following holds:

(10) α did not form an intention to see to it that q.

Since (7) implies

(11) α decided to see to it that q,

it becomes evident that the implication from (1) to (2) is not valid. Therefore, also in the case of weighing evidence for and against p, the argument from intention fails to show that (9) logically implies:

(12) α formed an intention to see to it that α forms the belief that p.

Our discussion of the argument from intention has revealed that a theory of agency should not make the notion of agency dependent on the notion of an intention.[7] *Stit*-theory is a formal theory meeting this requirement. In Section 4 we shall make use of this theory.

2.2. *The Argument from Missing Practical Reasons*

Further unfolding ideas of Pojman (1985), Audi (2001) has presented an argument from missing practical reasons. In the final analysis, this is an argument from indirectness: it is claimed that there is no *direct* voluntary control over beliefs. The argument aims at exploiting the distinction between theoretical reasons for beliefs and practical reasons for actions. According to Audi:

> There is causing oneself to believe something, and there is simply coming to believe, in the sense of a belief's forming, as where one sees the approaching tornado. Neither of these is an action of belief formation. Causing oneself to believe is a type of action that entails belief formation, but the belief formation is not a further action, any more than the raising of one's arm is an action beyond one's raising it. Here the belief formation is the *result* of an action, not, as the voluntarist would have it, something done at will. Moreover, whereas the arm's rising is the result of something done at will, the belief formation is at least normally the result only of something done by doing something else (2001, p. 101).

Thus, if an agent α causes herself to believe something, for Audi, α exercises only indirect control over her beliefs. The agent may perform some action, like contemplating a proposition. "This action will *cause* belief formation but is not an *act* of belief formation" (Audi, 2001, p. 104). Audi also compares raising one's arm with the arm's rising, caused by using a pulley. Whereas the former is said to be a directly willed action, the latter is said to be an indirectly willed nonaction. No matter how belief formation is "produced by the will," the result of the production is not an action. But why is forming a belief not an action? Audi assumes that "for any action, there at least can be a sensible expression of a practical reason" (Audi, 2001, p. 102) and such a reason "is

always expressible in a phrase of the form 'in order to *A*,' where '*A*' ranges over action-types'' (Audi, 2001, p. 99). Taking up an example used by William James, Audi explains that ''causing oneself to be more confident of jumping across a precipice'' does not express a practical reason for an action of forming the belief that *p*. In this case, *p* should be the sentence ''I can jump across a precipice.'' He maintains that ''although this expresses a reason for causing oneself to believe *p*, it does not in any clear sense express a reason for forming the belief that *p* if that is any different'' (Audi, 2001, p. 102). According to Audi, it is thus possible that I cause myself to believe that I can jump across a precipice in order to cause myself to be more confident of jumping across a precipice, but I cannot form the belief that I can jump across a precipice in order to cause myself to be more confident of jumping across a precipice, because belief formation does not admit of practical reasons.

First, I am not convinced that this discussion does justice to James's example. The idea is that I decide to believe that I can jump across a precipice in order to be able to jump across a precipice.

Second, I am not convinced that there are no practical reasons for deciding to believe. For example, a patient may decide to believe her doctor's explanations in order to bring about a certain change in the world, namely in order to finally obtain an adequate understanding of her illness. If James is right, it may happen that an agent α decides to believe that β will fall in love with α, in order to make it happen that β will fall in love with α. Likewise, I am not convinced that there are no practical reasons for belief formation. If it is assumed that belief formation is not a kind of action, then one could maintain that *therefore*, belief formation does not admit of practical reasons. However, Audi claims that belief formation fails to be an action type, because there are no practical reasons for belief formation.

According to Audi, causing oneself to believe is a generic action, whereas belief formation is a kind of event produced by such an action. In my view, Audi failed to show that,

because of missing practical reasons, belief formation fails to be an action type. Moreover, and maybe even more importantly, Audi has not explained why belief formation should be conceived of as "distinct from causing oneself to believe" (Audi, 2001, p. 104). Clearly, believing is distinct from a belief's formation. Therefore, causing oneself to believe is *different* from causing a belief's formation. A belief's formation in Audi's sense is a nonaction and thus not responsive to practical reasons. But this does not show that causing oneself to believe fails to be a generic action of forming a belief.[8]

If the voluntarist is right, belief formation is an *action type* that may be performed. It should then be adequate to understand "α forms the belief that p" as

(13) α sees to it that α believes that p.

Assuming that "α sees to it that p" is equivalent with "α sees to it that α sees to it that p," (13) is equivalent with

(14) α sees to it that α forms the belief that p.

The upshot of this discussion is that in order to obtain a formal explication of belief formation understood as an action type, a modal logic of agency may be combined with a (modal) doxastic logic.

2.3. *The Phenomenological Argument*

The problem of directness of control over beliefs also shows up in Feldman's discussion of the cases he is ready to concede as cases of deciding to believe something. Feldman explains that he has nonbasic immediate voluntary control over whether the lights in his office are on or off. This means that by doing something else over which he has basic voluntary control (say, turning a switch), he can see to it that the lights in his office are on or off. Since "belief about whether the lights are on tracks the actual state of the lights almost perfectly" (2001, p. 81), he has nonbasic voluntary control over whether he believes hat the lights are on. However, Feldman points out that whereas

the example shows that an agent may have control over evidence for having a certain belief, the problem is whether the agent does have control over how she *responds to evidence*. In cases of deliberation about what to believe this may appear to be the case.

Audi does not deny that there is freedom to believe or not when an agent is exposed to conflicting pieces of evidence.[9] He comments that it differs from freedom to act. Whereas freedom to believe consists in the absence of any "compelling *reason* for one to believe the proposition or its negation," in the case of freedom to act "there is no compelling *cause* operating on one toward the first action or toward the second" (Audi, 2001, p. 103). Audi admits that in the case of equally good evidence on both sides an agent may cease to withhold her preferred belief, but this action is not understood as directly forming a belief. It is "at best one that causes a belief to form from ... antecedent inclinations toward it" (Audi, 2001, p. 103).

But since beliefs do not come into being at random, what role then does the evidence play in belief formation? Does the evidence always directly trigger a belief's formation, or is it at least in some cases the agent who decides to believe, sometimes supported by the available evidence, sometimes against the evidence and sometimes unaffected by or irrespective of any evidence? Louis Pojman (1985) has presented a so-called phenomenological argument against doxastic voluntarism. Basically, the idea is that an inspection of the phenomenon of the emergence of beliefs reveals that the world forces beliefs on doxastic subjects. The argument goes as follows (Pojman, 1985, p. 40):

1. Acquiring a belief is a happening in which the world forces itself upon a subject.
2. Happenings in which the world forces itself upon a subject are not things the subject does (i.e., are not basic acts) or chooses.
3. Therefore, acquiring a belief is not something a subject does (i.e., is not a basic act) or chooses.

Pojman argues that the controversial first premise is supported by an introspective analysis of different kinds of belief acquisition. The types of coming into existence of beliefs considered (classified as kinds of beliefs) are: 1. perceptual beliefs, 2. memory beliefs, 3. abstract and logical beliefs, 4. "theoretical beliefs, including scientific, religious, ideological, political and moral beliefs" (1985, p. 42), and 5. testimony beliefs. The result of the introspection is that "[t]he Phenomenological Argument shows that volitionalism is abnormal and bizarre" (1985, p. 45).

I shall not enter into a detailed analysis of Pojman's discussion of the mentioned types of beliefs and their acquisition. I would just like to point out that among William James's genuine options, which are, according to James, cases of deciding to believe, there are choices between moral hypotheses and that the development of nonclassical logic may call into question the view that "we do not choose to believe that the law of noncontradiction has universal application" (1985, p. 42). Critical replies to Pojman's phenomenological argument may comprise several elements. (i) Evidence need not always be conclusive. The available evidence may just be not decisive and may leave room for different conclusions. Empirical theories, for instance, are usually under-determined by the observational data. (ii) It may happen that pieces of evidence contradict each other or that evidence is misleading. There may be contradicting testimonies, and, of course, there may also be erroneous or deceiving testimony. (iii) Proof methods may be contentious. An intuitionist, for example, will not believe the conclusion of a proof using König's Lemma, whereas his fellow classical mathematician is happy to accept the conclusion. (iv) Agent's may distrust the evidence even if it is conclusive. It may happen that α and β both in fact see an approaching tornado but that α mistakenly takes this seeing to be an illusion. Agents may, of course, be aware of (i)–(iv).

In other words, it is far from clear that all beliefs of all agents come into being as an inescapable response to some evidence. One might, of course, assume that any two distinct

agents are never confronted with exactly the same body of evidence just because of their different history, possibly different education, possibly very different social background etc. Also, different reactions of distinct agents to the same evidence can be explained without appeal to decisions to believe. The same holds for belief changes of single agents against a background of constant evidence. But still, it seems that an agent not only has options in evaluating the available information, but also in her response to the evidence.[10]

Although the phenomenological argument is far from being decisive, it leads to interesting questions. Are there any propositions a doxastic subject cannot but believe? If so, and if p is (a sentence expressing) such a proposition, then it should be impossible to decide to believe that p. Are there any propositions a doxastic subject cannot but fail to believe? If so, and if p is (a sentence expressing) such a proposition, then it should be impossible to decide to believe that p. The semantics of standard doxastic (epistemic) logic is such that "α believes (knows) that p" cannot be false if p is a tautology and cannot be true, if p is a contradictory formula. This is a strong idealization, but in any case, as a result, a semantics for ascriptions of decisions to believe that builds on standard doxastic (epistemic) logic ought to have it that the agentive sentence "α forms the belief that p" is unsatisfiable if p is a tautology or a contradictory formula.

3. THE SEMANTICS OF ASCRIPTIONS OF BELIEF FORMATION

Standard doxastic logic has been occupied chiefly with providing a semantics and proof theory for expressions of the shape "α believes that p." In considering belief formation and reading "α forms the belief that p" or, as we from now on also shall say "α voluntarily acquires the belief that p," as "α sees to it that α believes that p," the interest is shifted to more dynamic aspects of the central epistemological notion of belief.[11]

To obtain a semantics for the notion of voluntarily acquiring a belief, we may combine the semantics of the *dstit*-opera-

tor from *stit*-theory and the semantics of the standard belief operator from doxastic logic. The model theory for ascriptions of concrete actions in *stit*-theory uses Prior-Thomason branching time structures $\langle T, \leq \rangle$, see (Belnap et al., 2001) and references therein. The basic idea is that the moments of time collected in T have a tree-like structure. Such a tree branches toward the future, and this reflects the openness or indeterminacy of the future. There is no backward branching, and this condition is meant to capture the determinacy of the past. Maximal linearly ordered sets of moments from a tree are called histories. Intuitively, these histories may be seen as complete possible temporal developments of the world. If one considers a moment that belongs to a history, one may also say that this history passes through the moment, or that the moment occurs in the history. Since it is assumed that the future is open, branching may occur and more than one history may pass through a given moment. These ideas can be nicely depicted, see Figure 1.

In this setting, the indices at which sentences are interpreted are structured entities. A sentence is not just true or

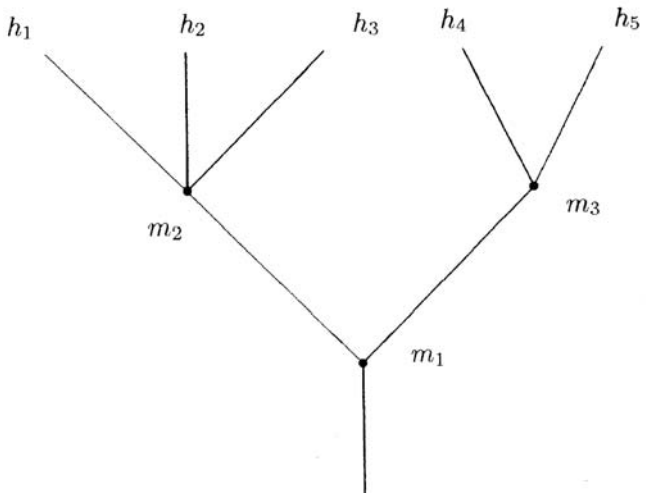

Figure 1. A branching tree of moments of time.

false at a moment but true or false at a moment/history-pair (m, h). The semantics of future contingencies makes this clear. Suppose that in Figure 1 A is true at moment m_3 and false at moments m_1 and m_2. What can be said about the truth or falsity of "Sometimes in the future A" at m_1? With respect to histories h_4 and h_5 the claim is true but with respect to histories h_1-h_3 it is not. Thus, future contingencies are evaluated at moment/history-pairs, and the idea of semantical equal treatment leads to evaluating all sentences at moment/history pairs.

Now we have to consider doxastic subjects. Each doxastic subject is supposed to be an agent who by her or his actions can influence the future course of the world. In *stit*-theory this idea is accounted for by assuming that for every individual agent, the histories passing through a moment are partitioned into sets of histories choice-equivalent for the agent. If two histories h and h' are choice-equivalent for an agent α at moment m, then α cannot discriminate by her or his actions at m between h or h'. The sets of histories choice-equivalent for an agent at a moment m represent the "choice-cells" of the agent at m. A very natural postulate then is that for every agent α, histories that pass through a moment m and divide only at a later moment must be choice-equivalent for α at m. The idea of choice-cells can also be nicely graphically represented, see Figure 2. Intentions of agents are so far *not* formally represented in this model. This is fine, because in Section 1 we have seen that there are unintentional actions and choices, and therefore intentions should not be built into the basic semantic picture. Recently, Müller (2004) has suggested introducing strategies as an additional parameter of semantic evaluation. Müller holds that a strategy may function "as an interface between our everyday mentalistic vocabulary that describes an agent's plans and intentions and the formal branching time framework" (2005, p. 202).

The *dstit*-theory developed by von Kutschera (1986) and Horty (1989) provides a simple semantics for ascriptions of concrete actions. The informally explained ideas can be turned into formal definitions.

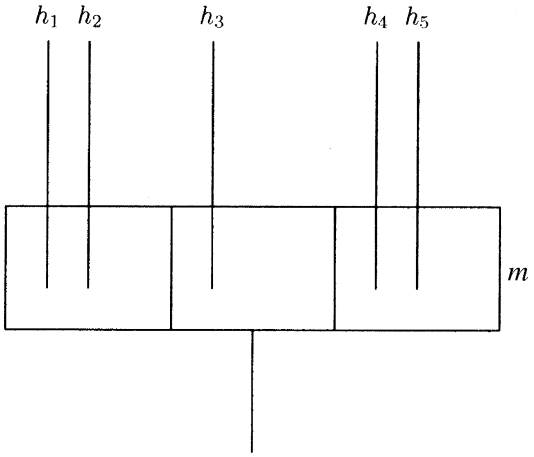

Figure 2. Choice-cells of an agent at moment m.

A pair $\langle T, \leq \rangle$, is called a *branching temporal frame* if T is a nonempty set (of moments), and $\leq$ is a partial order on T satisfying *historical connectedness* $(\forall m_1 \forall m_2 \exists m(m \leq m_1 \wedge m \leq m_2))$ and *no backward branching* $(\forall m \forall m_1 \forall m_2 ((m_1 \leq m \wedge m_2 \leq m) \supset (m_1 \leq m_2 \vee m_2 \leq m_1)))$. A *history in T* is a maximal set of moments (in T) linearly ordered by $<$, where $m < m'$ iff $m \leq m'$ and $m \neq m'$. The set of histories passing through moment m, H_m, is defined as $\{h \mid h$ is a history and $m \in h\}$. If $\langle \mathrm{T}, \leq \rangle$ is a branching temporal frame, then $\langle \mathrm{T}, \leq, Agent, Choice \rangle$ is called a *dstit frame*, if *Agent* is a nonempty set (of agents) and *Choice* is a function mapping every agent/moment-pair (α, m) to a partition of H_m (the histories *choice-equivalent for* α *at* m) satisfying *no choice between undivided histories* $(\forall \alpha \in Agent)(\forall H \in Choice(\alpha, m)) \forall h \forall h' [(h \in H \wedge \exists m'(m < m' \wedge m' \in h \cap h')) \supset h' \in H]$. If $h \in H_m$, then $Choice_{\alpha}^{m}(h)$ is the particular choice in $Choice(\alpha, m)$ containing h. A *dstit model* is a structure $\langle T, \leq, Agent, Choice, v \rangle$, where $\langle T, \leq, Agent, Choice \rangle$ is a dstit frame, and v is a valuation function that interprets atomic formulas by sets of moment/history-pairs. A complete axiomatization of dstit logic has been presented in (Xu 1998). The truth definition for [α *dstit*: A] ("α deliberately sees to it that A") is as follows:

DEFINITION 1. [α *dstit*: A] *is true in* $\langle T, \leq, Agent, Choice, v\rangle$ *at* (m,h)*iff*$(i)\forall h' \in Choice^m_\alpha(h)A$ *is true at* (m,h'), *and (ii)* $\exists h' \in H_m$ *such that* A *is not true at* (m,h').

A formula is said to be satisfiable in this semantics iff there is a dstit model $\langle T, \leq, Agent, Choice, v\rangle$ and a moment history pair (m,h) in this model at which A is true. The first condition in the above definition is called the positive condition; the second conditions is called the negative condition. The negative condition in a sense guarantees genuine choices. If A is a tautology or a contradictory formula, then [α *dstit*: A] is unsatisfiable.

We now build on the insight from Section 2.2 and combine the dstit semantics with standard doxastic (epistemic) logic, see (Fagin et al., 1995). The propositional logic of implicit belief (knowledge), the belief (knowledge) of completely rational, logically omniscient agents, is generally taken to be the polymodal logic KD45 (KT45 alias S5). For every agent α, "α implicitly believes (knows) that A" may thus be expressed as $B_\alpha A$ ($K_\alpha A$), where B_α (K_α) is a KD45 (S5) necessity operator. Intuitively, $B_\alpha A$ ($K_\alpha A$) is true at a moment/history-pair (m,h) if and only if A is true at every doxastic (epistemic) alternative for α at (m,h), that is, at every moment/history-pair compatible with what α believes (knows) at (m,h). Such ascriptions of implicit belief (knowledge) are nonagentive, because the claim that A is true at every state compatible with what α believes (knows) at (m,h) fails to describe a concrete action.

In order to interpret "α voluntarily acquires the (implicit) belief (knowledge) that A" (or "α forms the (implicit) belief (knowledge) that A") as [α *dstit*: $B_\alpha A$] ([α *dstit*: $K_\alpha A$)] the dstit models must be augmented by a doxastic (epistemic) accessibility relation between moment/history-pairs. A *doxastic (epistemic) dstit model* is a structure $\langle T, \leq, Agent, Choice, R, v\rangle$, where $\langle T, \leq, Agent, Choice, v\rangle$ is a dstit model and $R = \{R^m_\alpha | \alpha \in \text{Agent}, m \in T, R^m_\alpha \subseteq H_m \times H_m\}$ is a set of serial (reflexive), transitive and Euclidean relations. We then obtain the following truth definition for [α *dstit*: $B_\alpha A$] ([α *dstit*: $K_\alpha A$]) ("α deliberately sees to it that α implicitly believes (knows) that A"):

DEFINITION 2. $[\alpha\ dstit{:}B_\alpha A]([\alpha\ dstit : K_\alpha A])$ *is true in the doxastic (epistemic) dstit model* $\langle T, \leq, Agent, Choice, R, v\rangle$ *at* (m,h) *iff (i)* $\forall h' \in Choice^m_\alpha(h)$ $\forall h'' \in H_m$, *if* $(m,h')R^m_\alpha(m,h'')$ *then A is true at* (m,h''), *and (ii)* $\exists h', h'' \in H_m$ *such that* $(m,h')R^m_\alpha(m,h'')$ *and A is not true at* (m,h'').

In Figure 3, the moment m is partitioned into three choice cells for α. Moreover, the R^m_α alternatives to the histories passing through m are depicted by annotated arrows. The formula A is true at the moment/history-pairs (m, h_1), (m, h_3) and (m, h_4). In this simple example, at (m, h_2) it is true that α sees to it that α implicitly believes that A.

Note that the second, negative condition in Definition 2 prevents the formation of implicit belief (knowledge) from closure under logical consequence: if an agent voluntarily acquires the belief (knowledge) that A, and A logically implies B, then it does not follow that the agent also acquires the belief (knowledge) that B. Since for every logical omniscient agent α, $B_\alpha\top(K_\alpha\top)$ is valid, no agent can *decide* to implicitly believe (know) a valid formula $\top$; valid formulas are already implicitly believed (known). Since for every agent α and every moment m, the relation R^m_α is serial, $B_\alpha\bot(K_\alpha\bot)$ is

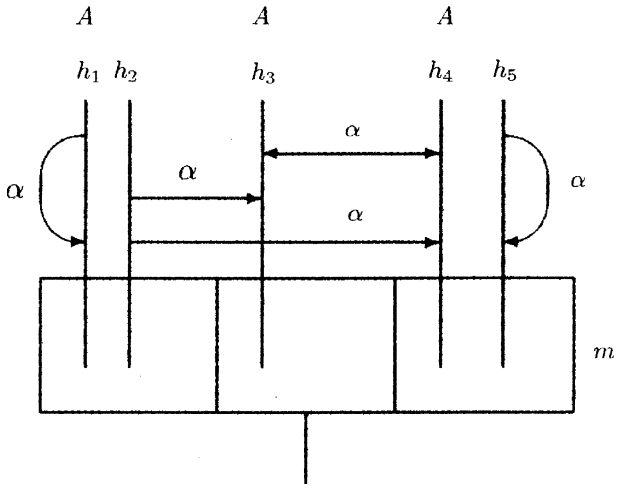

Figure 3. An example illustrating Definition 2.

false at every moment/history-pair and hence no agent can decide to implicitly believe (know) a falsehood $\bot$.

If not only a single doxastic (epistemic) subject but a group of agents Δ is considered, we need a semantics for ascriptions of collective voluntary belief (knowledge) acquisition:

($*$) Δ sees to it that Δ implicitly believes (knows) that A.

The clause "Δ implicitly believes (knows) that A" is unproblematic, since implicit, logically omniscient group belief (knowledge) is normally understood as implicit belief (knowledge) of all group members: $B_\Delta A (K_\Delta A)$ if and only if for every $\alpha \in \Delta, B_\alpha A$ $(K_\alpha A)$. Therefore, a semantics for $[\Delta \ \ dstit : B_\Delta A]([\Delta \ \ dstit : K_\Delta A)]$ ("Δ voluntarily acquires the implicit belief (knowledge) that A") is available if a semantics for $[\Delta \ \ dstit : B_\alpha A]([\Delta \ dstit : K_\alpha A)]$ is available. The latter is, however, straightforward, because there already exists a formal semantics for $[\Delta \ \ dstit : A]$. If $\Gamma \subseteq Agent$ and h is a history passing through $m \in T$, the set $Choice^m_\Gamma(h)$ of histories choice-equivalent with h for Γ at moment m is defined as $\{h' \mid (\forall x \in \Gamma) h' \in Choice^m_x(h)\}$.

DEFINITION 3. $[\Gamma \ \ dstit : B_\beta A]([\Gamma \ dstit : K_\beta A)]$ *is true in the doxastic (epistemic) dstit model* $\langle T, \leq, Agent, Choice, R, v\rangle$ *at* (m,h) *iff (i)* $\forall h' \in Choice^m_\Gamma(h), \forall h'' \in H_m$, if $(m, h') R^m_\beta (m, h'')$ *then A is true at* (m, h''), *and (ii)* $\exists h', h'' \in H_m$ *such that* $(m, h') R^m_\beta (m, h'')$ *and A is not true at* (m,h'').

We may observe that the formation of implicit group belief (knowledge) is neither closed under logical consequence nor closed under group membership. In other words, if $[\Gamma \ dstit : B_\beta A]([\Gamma \ dstit : K_\beta A)]$ and $\alpha \in \Gamma$, it does not follow that $[\alpha \ dstit : B_\beta A]([\alpha \ dstit : K_\beta A)]$ Let s_m be any mapping from *Agent* into the powerset of H_m such that $s_m(\alpha) \in Choice(\alpha, m)$. It has been suggested in *stit*-theory that the independence of agents can be captured by requiring that for every function s_m,

$$\bigcap\nolimits_{\alpha \in \mathrm{Agent}} s_m(\alpha) \neq \emptyset.$$

For a discussion of this plausible postulate see (Belnap et al., 2001, Section 10B). In Figure 4, moment m is divided into six cells, for $\Gamma = \{\alpha, \beta\}$, where the vertical lines separate β's choice cells C_1–C_3 and the horizontal line separates α's choice cells C'_1 and C'_2. The doxastic alternative relations of α and β are depicted by annotated arrows. In this example $[\Gamma dstit : B_\Gamma A]$ is true at moment/history-pair (m, h_2). The agents α and β do not act independently of each other. If α acts so as to realize h_4 and β acts so as to realize h_3, their joint action realizes *no* history.

4. THREE APPLICATIONS

The formal language and semantics of Section 3 may be seen as a useful apparatus for discussing in precise terms the disputed issues surrounding doxastic and epistemic voluntarism and the deontological conception of epistemic justification. In this final section, we shall briefly look at the concept of deciding to know, the idea of making believe, and a formal interpretation of the thesis of deontologism.

4.1. *Deciding to Know*

In the previous section, a formal semantics for ascriptions of both belief and *knowledge* acquisition has been defined. As a

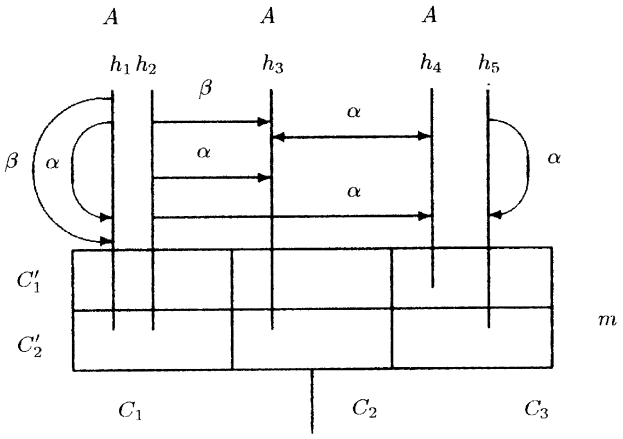

Figure 4. An example illustrating Definition 3.

first application of our analysis we shall discuss the concept of deciding to know implicitly. Initially, it might sound highly implausible that an agent may decide not only to believe implicitly but even to *know* implicitly. What could this possibly mean? If epistemic voluntarism is a species of doxastic voluntarism, since knowledge is a kind of belief, how should an agent decide to know something? The problem is that if α decides to know that p in the sense of in fact acquiring the knowledge that p, then α knows that p, and hence p is true. Can the epistemic subject in deciding to know guarantee the truth of what she decides to believe? Pojman (1985, p. 39) requires for doxastic voluntarism that "[t]he belief must be acquired independently of evidential considerations. That is, the evidence is not decisive in the belief formation." If α knows that p, then α in some way or another must be justified in believing that p. In the absence of "evidential considerations" this can only be some form of external justification. Can the agent "independently of evidential considerations" see to it that she is externally justified in what she decides to believe?

The semantics of epistemic dstit models helps clarifying these issues. An agent α decides to know that p at a moment/history-pair (m, h) iff at ever moment/history pair (m, h') such that h' is choice-equivalent with h for α at m, α knows that p at (m, h'). It is clear that "α sees to it that α knows that p" is satisfiable in some epistemic dstit model if p is neither a tautology nor a contradiction. Thus, if it is true at (m, h) that α sees to it that she knows that p, by her actions at (m, h) the agent can make sure that the future course of events comprises only histories h' such that p is true at every moment/history-pair (m, h'') compatible with what α knows at (m, h'). It is quite conceivable that agent α has the latter capacity. In Section 1 we have seen that there are unintentional actions. In an unintentional performance of a generic action, the agent need not be aware of performing that action. When voluntarily acquiring the implicit knowledge that p, the agent need not be aware of forming this implicit knowledge.

4.2. *Making Believe*

The second application of our analysis is realizing that an agent α can see to it that another agent β forms a certain belief only if β is not independent of α. This observation follows from the semantics of the *dstit*-operator. Let $\|A\|_m^{\mathcal{M}} := \{h \in H_m \mid A$ is true at (m,h) in model $\mathcal{M}\}$ and write $\mathcal{M},(m,h) \models A$ ($\mathcal{M}$, $(m,h) \not\models A$) if formula A is (not) true at (m, h) in $\mathcal{M}$.

Observation. For every dstit model $\mathcal{M} = \langle T, \leq, Agent, Choice, v\rangle$, for every moment/history-pair (m, h) from $\mathcal{M}$, for every α, $\beta \in Agent$ with $\alpha \neq \beta$, and for every formula A the following holds true: $\mathcal{M},(m,h) \not\models [\alpha\ dstit : [\beta\ dstit : A]]$.

Proof. By reductio. Suppose the claim is not true, and there exists a dstit model $\mathcal{M} = \langle Tree, <, Agent, Choice, v\rangle$, a moment/history-pair (m, h) from $\mathcal{M}$, agents $\alpha, \beta \in Agent$ with $\alpha \neq \beta$, and a formula A such that: $\mathcal{M},(m,h) \models [\alpha\ dstit : [\beta\ dstit : A]]$. Then:

(i) $Choice_\alpha^m(h) \subseteq \|[\beta\ dstit : A]\|_m^{\mathcal{M}}$ (positive condition) and
(ii) $H_m \neq \|[\beta\ dstit : A]\|_m^{\mathcal{M}}$ (negative condition).

If α and β are distinct agents, then any intersection of a choice cell for α with a choice cell for β in m is non-empty:

$$(*)(\forall K \in Choice_\alpha^m)(\forall K' \in Choice_\beta^m) K \cap K' \neq \emptyset.$$

Since $h \in Choice_\alpha^m(h)$, by (i) we obtain: $\mathcal{M},(m,h) \models [\beta\ dstit : A]$.
Therefore

(iii) $Choice_\beta^m(h) \subseteq \|A\|_m^{\mathcal{M}}$ (positive condition) and
(iv) $H_m \neq \|A\|_m^{\mathcal{M}}$ (negative condition).

By (iv) it follows that $(\exists h_1 \in H_m)(m,h_1) \not\models A$.
We now consider $Choice_\beta^m(h_1)$, the choice cell for β at moment m containing h_1. By (*) we have:

$$Choice_\alpha^m(h) \cap Choice_\beta^m(h_1) \neq \emptyset.$$

Let $h_2 \in Choice_\alpha^m(h) \cap Choice_\beta^m(h_1)$. Since $h_2 \in Choice_\alpha^m(h)$, we obtain by (i):

$$(**)(m, h_2) \models [\beta \; dstit : A]$$

and since $h_2 \in Choice^m_\beta(h_1)$ and $(m, h_1) \not\models A$, we have:

$$(m, h_2) \not\models [\beta \; dstit : A],$$

a contradiction with (**). ■

Thus, if α and β are distinct and mutually independent agents, it is logically impossible that α sees to it that β sees to something. And therefore, under this assumption about α and β, α cannot see to it that β forms a belief. Whereas $[\alpha \; dstit: B_\beta A]$ is satisfiable, $[\alpha \; dstit: [\beta \; dstit: B_\beta \; A]]$ is not. It follows that α is never responsible for β's acts of belief formation, but may be responsible for β's beliefs. Agent α might, for example see to it that β believes something by applying hypnosis or, perhaps, brain surgery.

Moreover, whereas making someone else, β, believe ($[\alpha \quad dstit: B_\beta A]$) is not equivalent to seeing to it that β forms a belief ($[\alpha \; dstit: [\beta \; dstit: B_\beta A]]$), we may note that making oneself believe

$$[\alpha \quad dstit : B_\alpha A]$$

and seeing to it that oneself forms a belief:

$$[\alpha \quad dstit : [\alpha \quad dstit : B_\alpha A]],$$

are equivalent.

4.3. *Deontologism*

We can finally return to the starting point of this paper: deontologism, the deontological conception of epistemic justification. The version of deontologism presented in Section 1 was:

> α is justified in believing that p if and only if α is permitted to form (or voluntarily acquire) the belief that p.

There is a familiar distinction between implicit and explicit permission. An agent α is implicitly permitted to see to it that p if α is not forbidden to see to it that p (or, equivalently, if α is not obligated to see to it that not p). Explicit permission

requires an act and not merely the absence of a prohibition. An even finer analysis is possible. An authority may refrain from forbidding (implicit permission) or, as it happens, just not forbid (nonprohibition). Accordingly, there are three versions of deontologism.

[1] α is justified in believing that p if and only if α is explicitly permitted to form the belief that p.
[2] α is justified in believing that p if and only if the norm-creating authority refrains from forbidding α to form the belief that p.
[3] α is justified in believing that p if and only if the norm-creating authority does not forbid α to form the belief that p.

In the previous section, we have taken one step in the direction of explicating the definiens of deontologism by combining *dstit*-theory and standard doxastic (epistemic) logic to obtain a semantically defined doxastic (epistemic) logic of agency. The next step is both obvious and viable: introducing deontic operators. One way of doing this is following the lines of (Horty, 2001) in order to formalize a utilitarian notion of personal obligation. Limitations of this approach have been worked out in (Murakami, 2005).

We conclude by pointing to the account of deontic operators in *stit*-theory suggested in (Wansing, 1998), which allows one to formally distinguish between explicit permission, implicit permission and nonprohibition. According to this proposal, "α is permitted to form (or voluntarily acquire) the implicit belief that p" means that there is an authority who permits α to voluntarily acquire the implicit belief that p. Let us abbreviate $[\alpha\ dstit\colon B_\alpha p]$ by $\alpha\ vab\ p$. We then obtain the following three readings of "β permits α to form (or voluntarily acquire) the implicit belief that p":

1. $[\beta\ dstit : (\alpha \quad vab \quad p \supset \neg S_\alpha)]$
   β explicitly permits α to form the belief that p
2. $[\beta \quad dstit : \neg[\beta \quad dstit : (\alpha \quad vab \quad p \supset S_\alpha)]]$
   β implicitly permits α to form the belief that p

3. $\neg[\beta \quad dstit : (\alpha \quad vab \quad p \supset S_\alpha)]$
β does not forbid α to form the belief that *p*

Here S_α is a propositional constant. Intuitively it stands for "there is wrongdoing of α." Thus "β forbids α to see to it that *p*" is translated into the formula [β *dstit*: ([α *dstit*: *p*] ⊃ *S*α)].

NOTES

[1] Pojman (1985, p. 37), for example, claims that "[i]t is widely held that we can obtain beliefs and withhold believing propositions directly by performing an act of the will" and points out that, among others, Thomas Aquinas, John Locke, William James, and Roderick Chisholm seem to have espoused doxastic voluntarism.

[2] Several other versions are possible, for a (partial) discussion see (Wansing, 2004).

[3] I assume that in the present context we do not exclude any relevant aspect from consideration, if the term "epistemically permissible" is replaced by "permissible." In other words, I shall, perhaps in a simplifying way, identify epistemic permission to believe with permission to believe.

[4] Actually, Feldman uses a concrete example:

> Bugliosi decided to write a book arguing that OJ is guilty.

No explanation is given why the central use of the notion of a decision is exemplified by a *retrospective* decision report in the past tense.

[5] In the literature one may find alternative understandings of the notion of a choice and the notion of an option. William James (1896), for example, defines an option to be a choice between two hypotheses.

[6] It has been objected by the anonymous referee of this paper that "one could simply claim that the *stit*-phrase in itself captures the decision-talk. Or one could take a more refined approach to the decision-action question and argue in a Davidsonian vein that the decision essentially involved in action need not operate on the same level of description as the *stit*-predicate." To a proponent of the first objection, I would reply that (8) does not imply (12) below. To a proponent of the second objection, I am inclined to reply that I am talking about actions and not about actions under certain descriptions. *Stit*-theory avoids quantification about concrete actions in its object-language. What comes closest to reifying actions in *stit*-theory is quantification about choice cells in the meta-language. Their identification does not depend on the use of (definite) descriptions.

[7] The anonymous referee of this paper is not convinced and argues that the above "sketchy example of an unintentional decision (deciding where to put my feet on the beach without intending it) is hardly convincing. e.g., in so far as I truly *decide* where to put my feet in a particular stride, this may essentially depend upon this stride's being part of an intentional action (produced e.g., by my intention of running down the beach in a certain way), even if there is no distinct intention operating on each stride of my running pattern. In general, one might claim with Davidson that an intention need only operate on one legitimate level of description of a certain event, in order for that to count as an action." My reply is that a decision *may* depend upon the stride's being part of an intentional action in some cases, but that this possibility does not prove that there is always such a dependence. In general, I believe that actions are not, in an ontological sense, dependent on descriptions.

[8] The verb "to form" (or at least its direct translation into my mother tongue) is ambiguous between an agentive and a nonagentive reading. Talking about belief acquisition or belief production seems to avoid such an ambiguity in favor of agentive readings.

[9] Both Feldman and Audi are talking about evidence. I should better use "information" here, if "evidence" is understood as information already interpreted by the doxastic subject.

[10] It may be objected that agents do not have direct voluntary control over their evidence evaluation to an extent that may count as direct voluntary control over their beliefs. This objection misses the point. Even if an agent lacks such a control and just passively finds herself confronted with inconclusive evidence concerning a proposition p, this does not by itself rule out that the agent (nevertheless) decides to believe that p.

[11] In this section, I am using material from (Wansing, 2002, 2004).

ACKNOWLEDGEMENTS

I would like to thank an anonymous referee for her/his critical and very constructive comments that allowed me to improve the presentation. The doxastic and epistemic responsibility for the beliefs expressed in this paper is, of course, taken by its author.

REFERENCES

Alston, W. (1988): 'The Deontological Conception of Epsitemic Justification', *Phlosophical Perspectives* 2, 257–299.

Audi, R. (2001): 'Doxastic Voluntarism and the Ethics of Belief', in M. Steup (ed.), *Knowledge, Truth, and Duty* (pp. 93–111), Oxford: Oxford UP.

Belnap, N., Perloff, M. and Xu, M. (2001): *Facing the Future: Agents and Choices in Our Indeterminist World*, New York: Oxford UP.
Engel, P. (1999): 'Volitionism and Voluntarism about Belief', in A. Meijers (ed.), *Belief, Cognition and the Will* (pp. 9–25), Tilburg: Tilburg UP.
Fagin, R., Halpern, Y., Moses, Y. and M., Vardi (1995): *Reasoning About Knowledge*, Cambridge/Mass: MIT Press.
Feldman, R. (2001): 'Voluntary Belief and Epsitemic Evaluation', in M. Steup (ed.), *Knowledge, Truth, and Duty* (pp. 77–92), Oxford: Oxford UP.
Ginet, C. (2001): 'Deciding to Believe', in M. Steup (ed.), *Knowledge, Truth, and Duty* (pp. 63–76), Oxford: Oxford UP.
Goldman, A. (1999): 'Internalism Exposed', *Journal of Philosophy* 96, 271–293.
Horty, J. (1989): *An Alternative Stit-operator*, Manuscript, Philosophy Department, University of Maryland.
Horty, J. (2001): *Agency and Deontic Logic*, New York: Oxford UP.
James, W. (1896): 'The Will to Believe', in *Essays in Pragmatism* (pp. 88–105), New York, Hafner Press, 1969.
von Kutschera, F. (1986): 'Bewirken', *Erkenntnis* 24, 253–281.
Lammenranta, M. (2004): 'Theories of Justification', in I. Niiniluoto, M. Sintonen and J. Woleński (eds.), *Handbook of Epistemology*. (pp. 467–497), Dordrecht: Kluwer Academic Publishers.
Müller, T. (2005): 'On the Formal Structure of Continuous Action', in R. Schmidt et al. (eds.), *Advances in Modal Logic*. Volume 5 (pp. 191–209), London: King's College Publications.
Murakami, Y. (2005): 'Utilitarian Deontic Logic', in R. Schmidt et al. (eds.), *Advances in Modal Logic* (pp. 211–230), London: King's College Publications.
Plantinga, A. (1993): *Warrant: The Current Debate*, New York: Oxford UP.
Pojman, L. (1985): 'Believing and Wiling', *Canadian Journal of Philosophy* 15, 37–55.
Ryan, S. (2003): 'Doxastic Compatibilism and the Ethics of Belief', *Philosophical Studies* 114, 47–79.
Shah, N. (2002): 'Clearing Space for Doxastic Voluntarism', *The Monist* 85, 436–445.
Steup, M. (2000): 'Epistemic Deontology and the Voluntariness of Belief', *Acta Analytica* 15, 25–56.
Wansing, H. (1998): 'Nested Deontic Modalities: Another View of Parking on Highways', *Erkenntnis* 49, 185–199.
Wansing, H. (2002): 'Seeing to it that an Agent Forms a Belief', *Logic and Logical Philosophy* 10, 185–197.
Wansing, H. (2004): 'Action-theoretic Aspects of Theory Choice', in S. Rahman et al. (eds.), *Logic, Epistemology and the Unity of Science* (pp. 419–435), Kluwer Academic Publishers, Dordrecht.

Xu, M. (1998): 'Axioms for Deliberative Stit', *Journal of Philosophical Logic* 43, 505–552.

Dresden University of Technology
Institute of Philosophy
01062, Dresden
Germany
E-mail: Heinrich.Wansing@tu-dresden.de