EDMUND HENDEN

INTENTIONS, ALL-OUT EVALUATIONS AND WEAKNESS OF THE WILL

ABSTRACT. The problem of weakness of the will is often thought to arise because of an assumption that freely, deliberately and intentionally doing something must correspond to the agent's positive evaluation of doing that thing. In contemporary philosophy, a very common response to the problem of weakness has been to adopt the view that free, deliberate action does not need to correspond to any positive evaluation at all. Much of the support for this view has come from the difficulties the denial of it has been thought to give rise to, both with respect to giving an account of weakness, as well as explaining the future-directed nature of intentions. In this paper I argue that most of these difficulties only arise for one particular version of the view that free, deliberate action must correspond to a positive evaluation, a version associated with Donald Davidson's account of weakness. However, another version of this view is possible, and I argue that it escapes the standard objections to the Davidsonian account.

1.

Many recent discussions of weakness of the will have taken as a point of departure Donald Davidson's proposed sceptical argument against the possibility of weakness.[1] In this paper I want to focus on a different sceptical argument and distinguish between two possible responses to it. What distinguishes them is that they have different views of intention, at the same time as they argue that this is the key to avoiding scepticism about weakness. While according to one view, intention is an unconditional all-out evaluative judgement, expressing the proposition that *doing A is better than doing B* (or any range of alternatives to doing A); according to the other, intention is a non-evaluative judgement, expressing the proposition that *A is to be done* or that *I shall do A*. Both views assume that intention is a form of acceptance of a practical conclusion that settles the question of what to do.

In what follows I shall examine what have been some of the standard objections to the evaluative view of intention and weakness of the will. Do these objections force us to abandon this view, as many of the critics appear to believe? I argue that the familiar objections only target *one specific version* of the evaluative view. I then propose a weaker version and argue

Erkenntnis **61:** 53–74, 2004.

that this version is not vulnerable to these objections. My conclusion is that there may be more to the evaluative view than critics have formerly tended to believe. Let me say something about the structure of this paper.

In the first half I present a sceptical argument against the possibility of weakness of the will, and distinguish between two possible responses to this argument, one claiming that intentions are all-out positive evaluations of a way of acting, the other that intentions are non-evaluative judgements that a certain action is to be done. I then present four standard objections to the evaluative view. In the second half of the paper I propose a weaker version of this view, and argue that it is not vulnerable to any of the standard objections. In the final section I consider and reject three possible objections to the weaker view.

2.

Its getting late and I realize that it is time for me to leave the party. I have an important day at work tomorrow and need to be rested. I should definitely not have another drink. But the party is fun. A friend of mine offers me another drink. I go through my reasons for and against one more time and judge that all-things considered I should go home. Without changing my mind, however, I accept the drink and stay for another hour.

This is a classic example of weakness of the will: the agent *freely*, *deliberately* and *intentionally* performs a particular action A against his judgement that some incompatible action B, would be better. In order to see why this description has seemed paradoxical to many, we need to formulate a sceptical argument against the possibility of weakness. There are different ways of doing this, the most well-known perhaps (at least in recent times), being Donald Davidson's in "How is Weakness of the Will Possible?".[2] Much has been written about Davidson's argument, and it seems fair to say that few have found it very convincing.[3] I shall not discuss it here; suffice it to say that much of the controversy surrounding it has been focused on *the concept of motivation* assumed; a widespread feeling has been that, contrary to what Davidson claims in this argument, our strongest motivation often part company with our judgement of what is the better course of action and, in fact, that weakness may be a very good example of a split of exactly this kind.[4] In what follows I want to focus on a different sceptical argument which does not depend on any particular claim about motivation. What is central to this argument is rather a particular view of *intention*.[5] Consider first the following claims about the connection between practical reasoning, intending and intentional action (note that when I talk about 'intending' and 'intentional action', I shall have in mind

only free, deliberate intentional action and freely and deliberately formed intentions):

(P1) S intentionally does A → S intends to do A.

(P2) S intends to do A = S accepts a practical conclusion that settles the question of what to do.

(P3) S accepts a practical conclusion that settles the question of what to do = S judges that doing A is better for S than the alternatives to doing A.

(P4) S intends to do A = S judges that doing A is better for S than the alternatives to doing A (from (P2) and (P3)).

(P5) S intentionally does A → S judges that doing A is better for S than the alternatives to doing A (from (P1) and (P4)).

(P6) S is weak-willed if and only if S freely, deliberately and intentionally does A against S's judgement that some incompatible action, B, is better than A.

The sceptical argument can now be stated as follows: suppose (P1)–(P6) are all true, and that S does A out of weakness. From (P6) it follows that S does A freely, deliberately and intentionally. Since S does A freely, deliberately and intentionally, it follows from (P5) that S judges that doing A is better for S than the alternatives to doing A. However, since S is weak-willed, it follows from (P6) that S also judges that doing B is better for S than doing A. So, if S is weak-willed it follows from (P1)–(P6) that S judges that doing A is better for S than doing B and that doing B is better for S than doing A. But just as it is impossible self-consciously to hold two formally inconsistent beliefs, it is impossible self-consciously to judge that each of two actions known to be incompatible is best to do. It follows that weakness of the will is impossible.

Let me say right away that space does not permit me to discuss all the claims and assumptions of this argument. My key concern in what follows will be with (P2), (P3) and, of course, (P4) which is entailed by them.[6] Very generally we can distinguish two approaches to the skeptical argument from authors who defend the possibility of weakness; we may call them *the evaluative* and *the non-evaluative* approaches. These are not the only possible approaches, but what is interesting about them, is that they both share certain basic assumptions regarding the connection between practical reasoning and intentional action; most notably they share the view that (P2) is true, that is, the view that intending is identical with accepting a practical conclusion that settles the question of what to do. The plausibility of this idea arise from the fact that forming intentions appears to settle the

question of what to do, and since accepting practical conclusions appears to do the same, it seems only natural to assume that forming intentions is *identical* with accepting practical conclusions. A further motivation for this view may be certain considerations having to do with *theoretical neatness*. Assuming (P2), we have a very nice analogy between theoretical and practical reasoning: just as one *believes* a proposition that is most probable on the basis of a consideration of one's evidence for believing, so one *intends* to perform an action on the basis of a consideration of one's reasons for acting.

The differences between the evaluative and non-evaluative approaches begin to appear when we consider the claim about practical conclusions in (P3). The non-evaluative theorists about intention argue that we need to abandon the view that accepting a practical conclusion about some action involves judging that performing that action is better than performing any alternative to it. The motivation is usually the view that the acceptance of (P3), given that we also accept (P2) and (P6), leads to a whole host of difficulties (scepticism about weakness is one; but there are many others), and that we therefore seem left with a choice between abandoning either (P2), (P3) or (P6), where the most reasonable alternative is to abandon (P3).[7] On the other hand, the motivation of the evaluative theorists to hold on to a version of (P3) is the traditional view that there must be *some* feature of what the agent freely and intentionally does that she finds *desirable*, and that the practical reasoning leading her to perform that action therefore must be *evaluative* reasoning; accordingly, what makes a piece of reasoning a piece of *practical* reasoning must be that it ensures the transmisssion of some *practical value* from premisses to conclusion, usually this value is thought of as 'goodness' or 'desirability'.[8]

So, given these differences between the evaluative and non-evaluative approaches, how do they seek to avoid scepticism about weakness? The answer, according to the non-evaluative theorists, is quite simple: if S does A out of weakness, she intends to do A, which means that she accepts a practical conclusion that *A is to be done* or *I shall do A*. There is no inconsistency between a conclusion with this content and a judgement that doing B is better than doing A. So, there is no conceptual difficulties in understanding how S can act against the latter judgement in a case of weakness. The non-evaluative theorists thus drive a wedge between the agent's evaluative judgement and her practical conclusion.

Since the evaluative theorists refuse to abandon the view that intentions are evaluative judgements, they have to find some other way to avoid scepticism. The most well-known approach, associated with Donald Davidson, is to drive a wedge between two kinds of evaluative judgements.[9]

On this view, the weak agent S is thought to make the conditional all-things considered judgement that doing B is better than doing A relative to the total set of relevant reasons available to her, but instead of moving on to make this judgement all-out or unconditionally, she makes the unconditional judgement that doing A is better than doing B, thereby violating *the principle of continence* which says that one should always do what one concludes is best on the basis of one's total set of available relevant reasons. Because the agent's all-things considered-judgement is "conditional" or *relativized* to the total set of relevant reasons available to her, while the judgement corresponding to her intentional action is "unconditional" or *non-relativized*, she is not entertaining a contradiction. On this view, in other words, we can hold on to all the claims in (P1)–(P6), given that we distinguish between conditional and unconditional evaluative judgements; while S's evaluative judgement in (P3)–(P5) is unconditional, the evaluative judgement she acts against if she does A out of weakness, is a conditional judgement. Let me now present what have been considered to be some main difficulties for the evaluative approach to intention and weakness of the will.

3.

There have been four main objections to the evaluative approach. While two of these have been directed at the proposed account of weakness, the other two have been directed at the account of intention. Let me start with the latter.

The first objection can be called *the uncertainty objection*.[10] A tram-driver loses control over the tram on a downhill stretch. He is able to steer the tram but cannot stop it. On approaching a fork in the line, he knows that there are two men working on one of the lines ahead of him, but he does not know which. Since he sees on the left a pile of bricks, and on the right two chalk marks on the track, he reasons that it is somewhat more likely that the men will be working on the line with bricks by the side of it and that he should therefore go right.

Now, in this case the agent is unsure what would be the better thing to do since he doesn't know which line the men will be working on. Given this uncertainty of the agent, it seems wrong to assume that he makes the all-out unconditional judgement that going right is better than going left. Still, he forms the intention to go right. The objection then, is that forming an intention in this case cannot be identical with making an evaluative judgement that going right is, unconditionally, better than going left. Instead it ought to be that going right is to be done.

The uncertainty objection clearly raises a difficulty for the evaluative view of intention. But this is not the only one. Another difficulty, pointed out by Michael Bratman, can be traced back to the role intentions play in coordinating our activities over time. Let us call it *the Buridan objection*.[11]

Suppose I know I can stop at one of two bookstores after work, Kepler's or Printer's Inc, but not both. However, I judge each alternative equally good or desirable, given my beliefs about the future. This seems to leave the evaluative theorist with a choice; either he can say that I intend to stop *both* at Kepler's and Printer's Inc since both are equally good or desirable, or he can say that I intend to stop at *neither* since one is not better than the other. Both alternatives, however, seem wrong. The first violates a very plausible *agglomerativity* constraint on rational intentions: if at one and the same time I rationally intend to do A and rationally intend to do B then it should be both possible and rational for me, at the same time, to intend to do A and to do B. Of course, since I know I cannot stop at both Kepler's and Printer's Inc, forming the intention to do both would violate this constraint. The second alternative also seems wrong. Clearly, I will decide to do *something* and, therefore, *intend* to do something. Suppose I decide to stop at Kepler's. But now the problem is that since I do not judge that stopping at Kepler's is better than stopping at Printer's Inc, my intention cannot be a comparative evaluative judgement in favour of stopping at Kepler's. Once again, we seem forced to conclude that intention cannot be identical with an all-out positive evaluation. And once again, the difficulty appears to be easily avoided if we adopt a version of the non-evaluative view.[12]

The two difficulties mentioned so far are difficulties for the evaluative view of intention. Let us now move on to some familiar difficulties which arise for the evaluative view of *weakness of the will*. One problem often mentioned, is that it does not escape the sceptical argument. Let us call this *the irrationality objection*. Consider again the following example: you are at a party and are trying to decide whether to have another drink or abstain. You go through your reasons for and against. In the end you judge that, all-things considered, abstaining is better than having another drink. But according to the evaluative theorists you conclude that having another drink *is better* than abstaining! In other words, you move from the premise that abstaining is better than having another drink *given the total set of relevant reasons available to you*, to the conclusion that having another drink is better than abstaining. But consider an analogous case of belief: suppose you believed that, *based on the total set of relevant evidence available to you*, it is more probable that smoking causes cancer than it is that it does not cause cancer. If you were to proceed from this premise to the

conclusion that it is more probable that smoking does not cause cancer, the chances are you would be accused of *lunacy* rather than irrationality![13] It simply does not seem plausible that the core cases of weakness should involve such extreme irrationality. In fact, one intuition we have is that the incontinent agent's failure is more a failure *to act* properly than *to reason* properly. What characterizes such agents is that they do not stick to their reasonable judgements, not that they mistakenly make unreasonable judgements.

The irrationality objection seems to be further evidence that the evaluative view is in trouble; once again the solution appears to be to adopt a version of the non-evaluative view. Let me finally mention one very common objection to the evaluative view of weakness.[14] The objection is that it is simply very *implausible* that the weak-willed agent judges that performing the incontinent act is *better* than performing the continent act. In general, we often seem to find ourselves in situations where we are more motivated to do one thing even though we judge that doing something else would be better. Why is not weakness of the will an example of such a case? Consider again the example where I freely and deliberately accept another drink against my own judgement that all things considered, it would be better to go home. According to the evaluative theorists, I conclude that drinking would be better than going home. But it just does not seem right, from a phenomenological point of view, to ascribe to me the judgement that drinking is *better* than going home. Suppose I were asked whether I thought it would be better to drink than to go home. If we assume that I am a clear eyed akrates, the chances are that I would sincerely deny this. However, the evaluative theorists must insist that I am wrong.

I now want to propose a different version of the evaluative view of intention and weakness of the will than the one commonly ascribed to Davidson, and argue that this version is neither vulnerable to the sceptical argument, nor to any of the objections mentioned above.

4.

Suppose we reject (P3), that is, the view that accepting a practical conclusion is identical with judging that doing A is better for S than the alternatives to doing A. One reason could be because we believe that the conclusion of practical evaluative reasoning is not itself an all-out evaluation, even if practical reasoning is evaluative reasoning.[15] Another reason could be because we believe that practical reasoning is not *evaluative* reasoning at all.[16] These have both been views adopted by non-evaluative theorists. However, I want to suggest a third possible reason, namely that

(P3) does not get *the form of the evaluative practical conclusion right*. In other words, while (P3) may be correct in claiming both that practical reasoning is evaluative reasoning and that practical conclusions are all-out evaluations, it may be wrong in claiming that practical conclusions are all-out evaluations of the form: "Doing A is better for S than the alternatives to doing A". One simple thought could be the following: the content of the practical conclusions accepted may vary from situation to situation depending on the specific context in which the judgement is formed. Which content is involved in a particular case could depend on factors such as the importance of the judgement for the reasoner, the time she has available for deliberation, her state of information, and so on.

Let me now propose the term 'action-worthy' as a general-purpose evaluative word to describe the content of practical judgements. That some action is judged to be *action-worthy* is for it to be seen by the agent as *worth doing* or *good enough* to perform. This judgement must be distinguished from the judgement that doing A is simply good or desirable since it seems possible to judge that doing A is good or desirable without judging it worth doing. Unlike the former judgement, the latter judgement has an inbuilt sufficiency condition. It is not only expressing the content that performing the action is seen by the agent as having *some* practical value; it is expressing the content that it is seen by her to have sufficient practical value to be done. We then have two different kinds of judgements, one *non-comparative*, the other *comparative*. Let me say a few words about each.

First, there are non-comparative judgements of the form, "Doing A is action-worthy"; the content of such a judgement involves no comparison of different alternatives. Judgements of this form may be seen to be examples of our tendency to satisfice, and may be based on a tiny subset of the relevant available reasons for and against doing A. Unlike the reasons that support comparative judgements, these reasons are not weighed against the reasons for and against any alternative to doing A. It is only the reasons for and against doing A as one option taken by itself that need be invoked. However, there may be cases in which the reasoner believes she lacks information to judge that doing A is action-worthy. Examples could be cases in which she believes that her set of reasons *fails* to support an inference to the conclusion that doing A is action-worthy, either because she believes that she has not considered all the relevant reasons she could obtain and ought to consider, or because she believes that, even if she has considered all these reasons, they are insufficient to conclude that doing A is action-worthy. If the reasoner believes her set of reasons insufficiently supports an

inference to the conclusion that doing A is action-worthy, she may settle for the less ambitious judgement that doing A *appears* action-worthy.

Second, there are judgements with a comparative content of the form, "Doing A is action-worthy (and doing B is not)". Even though judgements with this content imply that doing A is better than doing B, they must be distinguished from the latter kind of judgements, since it seems possible to judge that doing A is better than doing B without actually judging that doing A is *worth doing*. The judgement that "Doing A is action-worthy (and doing B is not)", is based on a comparison of the reasons for and against doing A with the reasons for and against doing B. How comprehensive this set of reasons is may vary from one situation to another. Sometimes the reasoner may want to consider the total set of available relevant reasons for and against doing A and doing B, sometimes she may be content with considering the reasons which have already occurred to her without necessarily believing that these are the total set of available relevant reasons. Once again, if the reasoner believes that her set of reasons insufficiently supports an inference to the conclusion that doing A is action-worthy (and doing B is not), she may settle for the judgement that doing A *appears* action-worthy (and doing B does not). Let us now first consider a case of *continent* practical reasoning.

Let 'R' denote the biggest set of available relevant reasons S has considered, and let 'A' be an action that S thinks is open to her. According to the view under consideration, S may make a conditional judgement of the form:

(1) Considering R, doing A is action-worthy.

The parallel in theoretical reasoning to (1), if we assume that 'E' is the biggest set of available relevant evidence S has considered and 'P' is a proposition, might be the belief that:

(1′) Considering E, P is belief-worthy.

The belief that some proposition is belief-worthy is for it to be seen by the agent as worth believing. That means that P is seen by the agent as being more likely than not-P. Belief worthiness is the analogue in theoretical reasoning to action-worthiness in practical reasoning.

Now, unrestricted detachment of the conclusion that doing A is action-worthy from the reasons R is unwarranted because there might be some other set of reasons, R′, that supports the conclusion that doing A is not action-worthy. To rationally detach the conclusion that doing A is action-worthy from her reasons, S needs a rule of detachment. On the strong

evaluative view associated with Davidson's account, this rule is *the principle of continence*, which holds that one should always do what one concludes is the best thing to do on the basis of one's total set of available relevant reasons. Clearly, this principle does not licence detachment of non-maximizing conclusions based on less than the agent's total set of reasons. To rationally detach conclusions about action-worthiness, the agent needs another type of rule.[17] I suggest that what is needed is a *default rule*, roughly of the following form:

(D) It is rational to draw the practical conclusion that doing A is action-worthy from your available relevant reasons unless defeating considerations occur to you or you believe that there is a significant chance that such considerations would occur to you if you undertook an investigation that it is reasonable for you to undertake.[18]

If the 'unless'-clause is not triggered, and no defeating considerations occur to S and she does not believe that there is a significant chance that such considerations would occur to her if she undertook any further investigation that it would be reasonable for her to undertake, she may move directly from (1) to an all-out unconditional judgement of the form:

(2) Doing A is action-worthy.

A parallel principle to (D), call it (D′), may be seen to govern the detachment of conclusions in theoretical reasoning. If the 'unless'-clause in (D′) is not triggered, S may move directly from (1′) to a belief of the form:

(2′) P is belief-worthy.

Let us now adopt a version of the evaluative view according to which S's intention to do A is identical with S's all-out unconditional judgement that doing A is action-worthy. I shall call this *the weak evaluative view* to distinguish it from the stronger view, associated with Davidson's account, that I described in Section 2.[19] So, to give an illustration, you may conclude that, considering that you are having a good time at the party and feel like another drink, having another drink would be action-worthy. If no defeating considerations occur to you, you simply detach this content, in accordance with (D), which then becomes the content of your intention. The theoretical parallel is that you conclude that, considering that the sky is red tonight it is belief-worthy that it will be sunny tomorrow. If no defeating consideration occur to you, you simply detach the latter content, in accordance with (D′), which then becomes the content of your belief.

So far, we have seen an example of a piece of continent practical reasoning resulting in a continent intention. But what would be an example,

according to the weak evaluative view, of a piece of *incontinent* practical reasoning, resulting in an *incontinent* intention? Let 'R' as before represent the biggest set of available relevant reasons S has considered and let 'r' be a subset of R. The weak-willed agent makes conditional judgements of the form:

(3) Considering r, doing B is action-worthy.

(4) Considering R, doing A is action-worthy (and doing B is not).

The parallel in theoretical reasoning to (3) and (4) are the beliefs that:

(3′) Considering e, P is belief-worthy.

(4′) Considering E, Q is belief-worthy (and P is not).

where 'E' is the biggest set of available relevant evidence S has considered and 'e' is a subset of E. Given (D), S ought to move to an all-out unconditional judgement of the form:

(5) Doing A is action-worthy (and doing B is not),

since R is the biggest set of reasons that S has considered, and the members of r are contained in R. Instead, however, S restricts her view to r and moves to an all-out unconditional judgement of the form:

(5)* Doing B is action-worthy,

thereby forming an intention with this content. In so doing, S is violating (D). This is because the conclusion in (3), from which the conclusion in (5)* has been detached, have been defeated by the conclusion in (4), which includes the biggest set of reasons S has considered. The parallel in the theoretical case is that, instead of moving to the belief that:

(5′) Q is belief-worthy (and P is not),

S moves to the belief that:

(5′)* P is belief-worthy,

thereby violating (D′) since (5′)* has been defeated by S's evidence in E. Let me give an illustration: you may judge that, considering that you are having a good time at the party and feel like another drink, having another drink is action-worthy. However, you may also judge that, considering that you feel like another drink and that you have important work to do

tomorrow, abstaining is action-worthy (and drinking is not). If no further defeating considerations occur to you, you should conclude, in accordance with (D), that abstaining is action-worthy (and drinking is not). Instead, however, you conclude that drinking is action-worthy, thereby violating (D) since this conclusion has been defeated by your reasons for abstaining. You are thus exhibiting weakness of the will.

Now, the weak evaluative view shares three key features with the strong evaluative view. First, both views see the step from premisses to conclusion in practical reasoning as *defeasible*. Second, they treat practical reasoning as *evaluative* reasoning. Third, they claim intentions are *all-out evaluative conclusions* of practical reasoning. However, unlike the stronger version, the weak view neither requires that agents only detach comparative conclusions supported by the total set of available relevant reasons or that they violate a principle of continence in cases of weakness.[20] In the next section I shall argue that these differences are sufficient to rescue the evaluative view from the objections of Section 3.

5.

The uncertainty objection claimed that intentions are not identical with unconditional judgements about what it is best to do, since sometimes an agent may form an intention to act even if he is unsure what is the best thing to do. This objection can be quickly passed over if we adopt the weak view. In the example of the tramdriver who is losing control over the tram, it is clear that since he is aware of lacking information about the consequences of his actions, his judgement will be that going right *appears* action-worthy (and that going left does not). This conclusion, unlike the one to the effect that going right is best, is compatible with the tramdriver's uncertainty as to whether going right *is*, unconditionally, the best option. Let us move on to the Buridan objection.

The Buridan objection may seem to pose a much bigger threat to the weak view. Consider again the case where I know I can stop at one of two bookstores after work, Kepler's or Printer's Inc, but not both and I judge each option equally desirable, given my beliefs about the future. The problem for the weak version can be set out as follows. Suppose I accept the following conclusions:

(6) Considering R, stopping at Kepler's is action-worthy.

(7) Considering R, stopping at Printer's Inc is action-worthy.

where 'R' is the biggest set of available relevant reasons I have considered. The difficulty is that, given (6) and (7), it seems rational for me to conclude:

(8) Stopping at Kepler's is action-worthy

and equally rational for me to conclude:

(9) Stopping at Printer's Inc is action-worthy

and, therefore, rational for me to conclude:

(10) Stopping at Kepler's is action-worthy & stopping at Printer's Inc is action-worthy.

But it is not rational for me to *intend* to stop at Kepler's *and* at Printer's Inc, since I know I cannot do both. At this point, the evaluative theorist needs to show that it is not rational of me to draw the conclusions in (8) and (9). One way in which this can be done serves to further clarify what it is to rationally form an intention according to the weak evaluative view.

On the weak view, a rational intention is not simply an evaluative judgement that doing A is action-worthy; it is an evaluative judgement of this form *that is governed by (D)*. (D) is a principle of rationality which tells the reasoner what kind of practical conclusion it is *rational* for her to draw (or, what kind of intention it is *rational* for her to form). It follows that there may be cases in which the reasoner draws a practical conclusion that doing A is action-worthy which does not proceed in line with (D). Examples could be cases of irrationality or non-rationality. In these cases the weak evaluative theorist will say that the agent does not form a *rational* intention to do A.

So, according to (D), under what conditions is it *not* rational for me to draw the practical conclusion that stopping at Kepler's is action-worthy? These would be cases in which this conclusion has been *defeated*. One example would be cases in which I believe that I have better reasons for concluding that some action other than stopping at Kepler's is worth doing. Another would be cases in which I believe that there is a significant chance that such considerations would occur to me if I undertook an investigation it is reasonable for me to undertake. A third example would be cases in which I believe that I have *equally good reasons* for drawing the conflicting practical conclusion that stopping at Printer's Inc is worth doing.

Why is the latter a *defeating* consideration? Simply because *rationality* cannot tell me which practical conclusion I ought to draw in this case.

That is, I have no way of telling that drawing one conclusion is the rational thing for me to do, just as I have no way of telling that drawing the opposite conclusion is the rational thing for me to do. Yet, the purpose of (D) is to tell me which practical conclusion *it is rational* for me to draw. In this kind of case, however, I realize that I cannot rely on (D) rationally to draw either conclusion since I know that neither of these conclusions would be legitimated by (D). So, since I cannot rationally draw either of these conclusions, it is not the case, on the weak view, that I can rationally intend to stop at Kepler's and rationally intend to stop at Printer's Inc. However, this does not rule out the possibility that I may *non-rationally plump* for one conclusion rather than the other. For example, I may non-rationally plump for the conclusion that stopping at Kepler's is action-worthy, thereby forming an intention with this content. This move is not governed by (D). Yet, because I have no better reasons for drawing one conclusion rather than the other, it is not *irrational*. It is simply non-rational: I am picking rather than choosing.

Now, if I am correct about the above, it may seem as if the weak evaluative theorist has a way of countering at least two of the objections to the evaluative view of intention I described in Section 3. But what about the objections to *the view of weakness* that seemed to flow from this view? It is reason to believe, I think, that the weak evaluative theorist can also avoid these objections. Let me start with the irrationality objection.

The irrationality objection claimed that a consequence of the evaluative view is that the incontinent agent is being represented as being *too irrational*. Is the weak version of this view vulnerable to this objection? Consider first the premisses of the incontinent agent's reasoning on the weak view compared with on the strong view. While on the weak view, these premises express the agent's evaluative assessment of *the biggest set* of available relevant reasons she has considered, on the strong view they express the agent's evaluative assessment of *the total set* of relevant reasons available to her. Since the biggest set of available relevant reasons the agent has considered need not be equivalent to the total set of relevant reasons that is available to her, the weak view does not imply (unlike the strong view) that the agent has *in fact* assessed the total set of relevant reasons that is available to her, or that she *believes* that she has assessed this total set. Her premisses are, therefore, potentially *weaker* than on the strong view.

What about the principles of rationality that the incontinent agent is supposed to violate in cases of weakness? While on the weak view, the agent is supposed to violate a *default rule* like (D), on the strong view she is supposed to violate the *principle of continence*. In order to violate the

principle of continence, which says that one should do what one concludes is best on the basis of all one's relevant available reasons, the agent must be assumed to *possess* this principle, that is, her reasoning must be governed by it. To reason in accordance with the principle of continence, she must check that her practical conclusion picks out the best of her options and is based on the total set of relevant reasons that is available to her. This requires that she rules out every relevant consideration contrary to the best of her options and also that she compares this option with alternative options which might also be open to her. To violate the principle of continence thus requires that she engages in an exhaustive piece of practical reasoning in which she *maximizes* both her available evidence and the practical value of her chosen act, and then knowingly draws the wrong conclusion.

Now, compare this with what is involved in violating a default rule like (D). To reason in accordance with (D), the agent may sidestep intermediate steps, such as checking that her reasons are *the total set of available relevant reasons* or making sure that it picks out *the best* of her options. Instead, she may legitimately detach her conclusion *directly* if no defeating considerations occur to her; she does not have to think that since no defeating considerations occur to her, she should draw the conclusion. Rather, whenever no such considerations come to mind, she simply draws the conclusion (this requires, of course, that she is sensitive to the defeasibility conditions of her reasoning!).[21] What this demonstrates is that default rules like (D) impose less rational constraints on the agent's reasoning than the principle of continence. To violate a rule like (D), the agent does not have to engage in any exhaustive piece of practical reasoning, as she has to do to violate the principle of continence; it suffices that she knowingly *jumps* to a defeated conclusion, whether the *un*defeated conclusion which she ignores is based on an assessment of just *a few* or *most* of her available relevant reasons. This suggests that, rationally speaking, it *may* be easier to violate a default rule and the piece of reasoning it governs, than it is to violate the principle of continence and the piece of reasoning it governs (more on this below).

Consider now, finally, the conclusion of the incontinent agent's reasoning on the weak view compared with on the strong view. While on the weak view, this conclusion is expressing the *non-comparative* content that performing the incontinent action is *worth doing*, on the strong view it is expressing the *comparative* content that performing the incontinent action is *better* than performing the continent action. The latter content is clearly stronger than the former since it is possible to judge something worth doing without necessarily judging it better than all other alternatives.

To summarize, on the weak evaluative view, the incontinent agent reasons from potentially weaker *premises* than on the strong evaluative view, her violation of the principle of rationality is potentially *less severe* than on the strong evaluative view and, finally, the content of *her incontinent conclusion is in fact weaker* than on the strong evaluative view. Together this suggest that the weak evaluative view allows the agent more *latitude* to draw the incontinent conclusion than the strong view. An agent has latitude to draw an incontinent conclusion if her premises do not actually entail the continent conclusion. Latitude is, however, a matter of degree.[22] Thus the agent's latitude will *diminish* as her premisses become stronger and make the continent conclusion better supported. Were the agent to have full access to *the complete set* of considerations relevant to the value of a certain action, her premisses would *logically entail* the continent conclusion, and her latitude to draw the incontinent conclusion would, as a consequence, vanish altogether. In that case she would be making a logical mistake if she went on to draw the incontinent conclusion.

Does this idea of 'more latitude' offer a way to answer the irrationality objection? The reason the answer must be yes, I think, is because more latitude diminishes the irrationality of the weak agent's practical reasoning. Why is that? Because the more latitude the agent has, the more *inconclusive* her evidence is and the easier it is for her to make *a wishful guess* at the practical value of performing the incontinent action. For example, consider once again the case where I accept another drink at the party although I judge that I should go home. In that very moment, when I am aware of forming the incontinent intention against my own better-judgement, am I not also thinking that perhaps only *one* more drink does not really matter after all? That perhaps I am exaggerating the effects on my performance at work tomorrow? That if I *really had* taken into account *the total set* of relevant reasons available to me, perhaps I would have concluded that having another drink would be worth doing? Such uncertainty, it seems, can co-exist with the judgement that it would be a mistake to have another drink given the biggest set of available relevant reasons I have considered, that is, it does not have to cause me to change my mind about what I should do. So, I can still recognize the irrationality of my own state of mind.

Let me finally address the last of the objections I mentioned in Section 3. This was the objection that it is simply very *implausible* that the weak-willed agent judges that performing the incontinent act is better than performing the continent act; on the contrary, it was claimed, it is much more plausible that she judges that *the continent act* is better, but fails to be motivated in accordance with this judgement. It is difficult to disagree with this objection. However, it is quite clear that it only threatens the strong

version of the evaluative view, but leaves the weak version unharmed. In contrast with the strong view, the weak view accepts that there may be *no respect* in which the incontinent action appears *better* than its continent alternative from the agent's perspective (that she ranks it higher on some scale of values). Still, it has some *intrinsic* practical value for her; for example, it may give her a certain *form of pleasure* that she believes that she would not get out of performing the continent action but that she ultimately values less than the pleasure she believes that she would get out of the latter. By restricting her view to these valuable features of her incontinent action, she detaches the conclusion that it is *worth doing* even though she knows that this conclusion has been defeated by the conclusion that is based on the larger set of reasons she has considered.

In the final section of this paper I shall consider three possible problems for the weak view.

6.

Let me begin by considering one objection that may have occurred to some readers. Since in order to rationally intend to do A, one must judge that doing A is at least *better* than not-doing A, it follows that rationally intending to do A cannot be identical with the *non-comparative* judgement that doing A is action worthy. If we are to identify intending with an evaluative judgement at all, it better be with a *comparative* evaluative judgement!

There is a simple answer to this objection. It is, of course, correct that if the reasoner thought that not-doing A was better than doing A, she could not rationally form an intention to do A. However, to legitimately detach her non-comparative conclusion that doing A is action-worthy it does not actually have to occur to her that doing A is better than not-doing A. It is sufficient that no thoughts *to the contrary* occur to her. In other words, her judgement that doing A is action-worthy depends on the *non-occurrence* of the thought that not-doing A is better than doing A. It does not have to be based on an evaluative comparison of doing A with not-doing A.

The second objection I shall consider appears slightly more serious. It may be argued that in theoretical reasoning we typically take a further step from the judgement that 'P is *belief-worthy*' to the conclusion that 'P is *true*'. By analogy, there should be in practical reasoning a further step from the judgement that 'Doing A is *action-worthy*' to the conclusion that 'A *is to be done*'. If this is correct, it suggests that the content of the intention should be 'A is to be done' rather than 'Doing A is action-worthy', just as the content of belief is 'P is true' rather than 'P is belief-worthy".[23]

The reply to this objection is that we need to distinguish between the level of reference and the level of description when we speak of the content of intentions and beliefs. On the level of reference, believing that P is belief-worthy and believing that P, are not two different mental states with a transition between them; they are the same state. For this reason they do not correspond to different stages in the agent's theoretical reasoning. Evidence that they are the same state is that you cannot be in a state of believing that P is belief-worthy and *not* be in a state of believing that P (this does not exclude cases of self-deception, since in these cases you do believe what you believe is belief-worthy. The trouble is that you also believe the negation of it). Of course, this does not rule out that you can *think* about this state in different ways, i.e. as simply the belief that P or the belief that P is *belief-worthy*. By analogy, it is in the nature of judging that doing A is action-worthy (and doing B is not), that you judge that A is to be done (and doing B is not). Since these judgements are different descriptions of one and the same state, they do not correspond to separate stages in the agent's practical reasoning. However, when *analyzing* practical reasoning and intentional action, the appropriate level of description may be in terms of practical value, i.e., *action-worthiness*.

Let me end by mentioning a possible objection against the account of weakness of the will proposed by the weak evaluative view. According to the weak view, the uncertainty created by latitude combines with the weak agent's wishful guess at the practical value of the incontinent act to make the transition to the irrational state easier. The objection would be to ask what it is about such a case that makes it different from a case in which the agent judges that there may be a reason not now available to her which will show that doing what now appears to be an incontinent action is in fact what she rationally ought to do. In both cases, the agent may recognize that her deliberation is less than perfectly complete, but in the latter case she would not necessarily be weak-willed. The answer is that it depends on whether the agent treats her judgement that there may be a reason not now available to her, *as a* defeating consideration or not. If she takes it to actually *defeat* her judgement that she should perform what now appears to be the continent action, then her case is not a case of weakness of the will, but rather a case of changing her mind about what would be action-worthy. On the other hand, if she does not treat it as a defeating consideration, then she will be weak-willed if she goes on to perform the incontinent action. Let me now summarize the conclusions of this discussion.

7. CONCLUSION

I have presented a sceptical argument against the possibility of weakness of the will and distinguished between two possible responses to it. These responses differed in their view of intention. While, according to both views, intentions are conclusions of practical reasoning, the evaluative theorists claimed these conclusions were evaluative, while the non-evaluative theorists maintained they were non-evaluative. One important set of considerations in favour of the latter view was the various problems the evaluative view seemed to run into. I mentioned four standard objections against the evaluative view from this perspective. Then I argued that these objections in fact target a *specific version* of the evaluative view. In support of this claim, I proposed a weaker version of the evaluative view and argued that this version is not vulnerable to these objections. If I am correct, it shows that there may be more resources in the evaluative view of intention and weakness of the will than previously believed. Whether an evaluative view in the end turns out to be the view we should adopt, is a further question that I have not addressed in this paper. The answer to that question will depend on features of the non-evaluative view, as well as on our view of practical reasoning in general.

NOTES

[1] I would like to thank Bill Child, David Charles, Richard Holton and John Broom for helpful comments on earlier drafts of this paper.

[2] Davidson, D.: 1980, "How is Weakness of the Will Possible?", *Essays on Actions & Events*, Oxford: Clarendon Press, pp. 21–43.

[3] Especially since Gary Watson's 1977 article. Watson, G.: 1977, "Scepticism about Weakness of the Will", *Philosophical Review* 86, pp. 316–339.

[4] This has been one of the most common objections to Davidson's own account of weakness. See for example Watson (1977), Taylor (1980), Mele (1987).

[5] M. Bratman discusses a similar argument in his 1979 article "Practical Reasoning and Weakness of the Will". However, the following formulation of the argument is mine. See Bratman, M.: 1979, "Practical Reasoning and Weakness of the Will", *Nous*, pp. 155–171.

[6] Obviously, questions can be raised, especially, about (P1), which Michael Bratman has called *the Simple View* and vigorously opposed (see Bratman, M.: 1999, "Two Faces of Intention", *The Philosophy of Action*, ed. Alfred R. Mele, Oxford University Press, pp. 15–34). I shall have to leave questions about the Simple View aside. However, it should be noted that the sceptical argument does not commit one to accepting the Simple View as *a general truth*. By restricting (P1) to cases of weakness of the will, the sceptical argument may be correct, even though the Simple view is false about some cases of intentional action.

[7] Authors who appear to accept some version of the non-evaluative view of intention, include Michael Bratman (1979), David Charles (1984), Paul Grice (1985), Hugh J. McCann (1998), Alfred Mele (1987), Christopher Peacocke (1985) and David Pears (1984).

[8] This involves some degree of legislation about the meaning of 'desirable'. In ordinary language, there might be other (more specific) ways of using the term. For example, we sometimes use the term to talk about the subset of all our reasons that has to do with personal preference: in that sense, we might say that, even though doing B was *more desirable* than doing A, doing A was nonetheless *the better* thing to do (because, for example, it was one's duty to do A). But no natural-language evaluative term is free from such ambiguity. If one believes some general-purpose evaluative word is needed to represent the nature of practical reasoning, 'desirable' appears to be as good a candidate as any. When I talk about 'desirability' or 'goodness' in what follows (I will not distinguish between them), it should be interpreted as broadly as possible; it is just meant to be a general-purpose evaluative word that shows how practical premisses support practical conclusions by representing them as giving evidence for these conclusions. A statement of the evaluative view of practical reasoning can be found in H. P. Grice's *Aspects of Reason*. Here Grice says: "I would regard reasoning as a faculty for enlarging our acceptance by the application of forms of transition, from a set of acceptance to further acceptance which are such as to ensure the transmission of value from premisses to conclusion [. . .] By 'value' I mean some property which is of value (of a certain *kind* of value, no doubt). Truth is one such property, but it may not be the only one; and we have now reached a point at which we can identify another, namely practical value (goodness)" (Grice, H. P. 2001: *Aspects of Reason*, ed. Richard Warner, Oxford: Clarendon Press, p. 87).

[9] See Davidson, D.: 1980, "How is Weakness of the Will Possible?".

[10] The following version is due to David Charles. See Charles, D.: 1983, "Rationality and Irrationality", *Proceedings of the Aristotelian Society*, p.197. A similar example can be found in Bratman, M.: 1979, "Practical Reasoning and Weakness of the Will", p. 161.

[11] Bratman, M.: 1985, "Davidson's Theory of Intention", *Actions and Events, Perspectives on the Philosophy of Donald Davidson*, eds. Lepore & McLaughlin, Oxford: Basil Blackwell, pp. 14–28.

[12] Note that it does not get the evaluative theorist off the hook to argue that in Buridan cases, we give ourselves a reason to prefer one alternative to the other, e.g., by flipping a coin. As Michael Bratman has pointed out, this only pushes the problem back; why assign 'stopping at Kepler's' to heads rather than 'stopping at Printer's Inc'? Ibid. p. 28.

[13] I owe this example to Hugh J. McCann. See McCann, H. J.: 1998, "Practical Rationality and Weakness of the Will", *The Works of Agency, On Human Action, Will, and Freedom*, Cornell University Press, p. 225.

[14] This objection is mentioned by Bratman (1979), Charles (1984), Pears (1984), Peacocke (1985), McCann (1998).

[15] This appears to be Michael Bratman's view. See Bratman, M.: 1979, "Practical Reasoning and Weakness of the Will".

[16] This appears to be David Charles's view. See Charles, D.: 1983, "Rationality and Irrationality".

[17] In fact, it is a strength of the weaker version that it does not depend on the principle of continence. As many critics have pointed out, the principle of continence may seem *too demanding*. As one author has put it, practical reasoning "does not require that I dredge up all of my reasons [. . .] puzzle out the best assessment I can of the relative advantages of each route, and reach a solemn judgement that in light of all my reasons, I-45 (say) stands as my best option". See McCann (1998), p. 221.

[18] How significant a chance has to be in order to make it irrational to draw the conclusion in favour of doing A may be context-dependent. The worse it would be to make a mistake,

the smaller the chance of finding a defeating condition would need to be in order to count as significant enough. Also, the 'investigation' in this formulation refers to an investigation that it is reasonable for you to undertake *given your present circumstances right now*. For the sake of convenience I shall stick to the simpler formulation in (D). There have been other authors who have suggested that practical reasoning may be governed by principles with this general structure. One example is Michael Bratman who in his 1979 article on weakness of the will proposes a principle of rationality similar to (D). However, the important difference between Bratman's principle and (D), is that while the role of Bratman's principle is to detach non-evaluative practical conclusions from evaluative premisses, the role of (D) is, in addition, to transfer practical value from evaluative premisses to evaluative practical conclusions. For an objection to Bratman's claim that evaluative practical reasoning can support non-evaluative practical conclusions, see Charles, D.: 1983, "Rationality and Irrationality".

[19] Let me just mention one worry some may have about the proposed representation of the agent's practical judgements on the weak view. The worry is that the nature of the agent's reasoning on this view, suggests that the agent's *reasons* need not be part of the content of her judgement at all. In other words, why can she not move directly to an unconditional all-out judgement with a content of the form: "Doing A is action-worthy"? The answer is that she can. It is in the nature of default practical reasoning that one may jump straight to one's conclusions. Perhaps in connection with routine actions such as writing your name, opening up a door or putting on your shoes, where you decide to act immediately without having to think how, no reasons need be part of the content of your judgement. But default practical reasoning is not limited to such cases; it may, in addition, be part of more elaborate cases of practical reasoning that require several steps. Such reasoning will include (some of) the agent's reasons as premisses. An illustration of the role of defaults in more elaborate cases of reasoning is provided by Kent Bach: "When our reasoning is sufficiently complex, we do not survey the entire argument for validity. We go more or less step by step, and as we proceed, we assume that if each step follows from what precedes, nothing has gone wrong. This is not always so, for an implausible conclusion along the way may lead us to question some previous step (either a premise or a bit of reasoning). An intermediate conclusion will seem implausible if it conflicts with other beliefs. Of course there is no guarantee that we will detect every such conflict, but we implicitly assume that when there is one, we will detect it and go back over our reasoning. Here we rely on our ability to detect such conflicts. Even if our lines of reasoning were always perspicuous, so that we could view them as a whole, there would still be points at which we do not actually check for validity but simply 'go along' with the reasoning at that point. We just 'see' that the next step follows" (Bach, K.: 1984, "Default Reasoning: Jumping to Conclusions and Knowing When to Think Twice", *Pacific Philosophical Quarterly* 65, pp. 37–58).

[20] I will not speculate as to what extent what Iehave called the weak version of the evaluative view preserves the spirit of Donald Davidsons's original proposal (which generally is thought to be equivalent to what I have called the strong version of the evaluative view), whether it can be seen as an expansion of his views, a modification or perhaps different in some decisive way. That being said, I cannot see any reason why Davidson should be committed to only one kind of evaluative predicate expressed by the two-place relation of 'betterness'. In general, the weak view preserves the main elements of Davidson's approach to intention and weakness of the will.

[21] Regarding the nature of default rules, I am indebted to Kent Bach's discussion (1984).

[22] Another philosopher who has emphasized the importance of the notion of latitude is David Pears. Pears argues that it is central to the understanding of Donald Davidson's

account of weakness of the will. See Pears, D.: 1982, "Motivated Irrationality', p. 163. If it is correct to ascribe something like the strong view to Davidson, it can be objected that he does not sufficiently exploit the resources the notion of latitude offers in order to preserve an evaluative framework for the understanding of intention and weakness of the will.

[23] Thanks to David Charles for forcing me to address this objection.

REFERENCES

Bach, L.: 1984, 'Default Reasoning: Jumping to Conclusions and Knowing When to Think Twice', *Pacific Philosophical Quarterly* **65**, 33–58.

Bratman, M.: 1979, 'Practical Reasoning and Weakness of the Will', *Nous* 155–171.

Bratman, M.: 1985, 'Davidson's Theory of Intention', in LePore and McLaughlin (eds.), *Actions and Events, Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, Oxford, 1428 pp.

Bratman, M.: 1999, 'Two Faces of Intention', in Alfred R. Mele (ed.), *The Philosophy of Action*, Oxford University Press, pp. 15–34.

Charles, D.: 1983, 'Rationality and Irrationality', *Proceedings of the Aristotelian Society*, pp. 191–212.

Davidson, D.: 1980, 'How Is Weakness of the Will Possible?', *Essays on Actions & Events*, Clarendon Press, Oxford, pp. 21–43.

Grice, H. P. and J. Baker: 1985, 'Davidson on Weakness of the Will', in Vermazen and Hintikka (eds.), *Essays on Davidson, Actions and Events*, Clarendon Press, Oxford, pp. 27–49.

Grice, H. P.: 2001, in Richard Warner (ed.), *Aspects of Reason*, Clarendon Press, Oxford.

McCann, J.: 1998, 'Practical Rationality and Weakness of the Will', *The Works of Agency, on Human Action, Will and Freedom*, Cornell University Press, pp. 213–233.

Mele, A.: 1987, *Irrationality, An Essay on Akrasia, Self-Deception, and Self-Control*, Oxford University Press.

Peacocke, C.: 1985: 'Intention and *Akrasia*', in Vermazen and Hintikka (eds.), *Essays on Davidson, Actions and Events*, Clarendon Press, Oxford, pp. 51–73.

Pears, D.: 1982, 'Motivated Irrationality', *Proceedings of the Aristotelian Society*, Vol. 56, pp. 155–177.

Pears, D.: 1984, *Motivated Irrationality*, Oxford University Press.

Taylor, C. C. W.: 1980, 'Plato, Hare and Davidson on Akrasia', *Mind* **LXXXIX**, 499–518.

Watson, G.: 1977, 'Scepticism About Weakness of the Will', *Philosophical Review* **86**, 316–339.

Department of Philosophy
University of Oslo
Blindern
0315 Oslo
Norway
E-mail: edmund.henden@filosofi.uio.no

Manuscript submitted 25 April 2003
Final version received 27 November 2003