



# Conflicting Judgments and Weakness of Will

Nora Heinzelmann<sup>1,2</sup> 

Received: 26 March 2020 / Revised: 16 July 2020 / Accepted: 29 July 2020 /  
Published online: 25 August 2020  
© The Author(s) 2020

## Abstract

This paper shows that our popular account of weakness of will is inconsistent with dilemmas. In dilemmas, agents judge that they ought to do one thing, that they ought to do something else, and that they cannot do both. They must act against either of their two judgments. But such action is commonly understood as weakness of will. An agent is weak-willed in doing something if she judges that she ought to and could do something else instead. Thus, it seems that, in a dilemma, the agent is weak-willed by definition. But this is puzzling: clearly, the two are different phenomena. The puzzle may support scepticism about weakness of will or dilemmas. Here, I argue that the two are consistent on a revised understanding of weakness of will. To do so, I further distinguish the mental states of an agent in a dilemma from those of a weak-willed person.

**Keywords** Dilemmas · Weakness of will · Conflict · Judgment · Davidson

## 1 Introduction

In a conflict of judgment, the agent must pick one of several<sup>1</sup> mutually exclusive options without being aware of decisive reasons to choose one rather than another. We can divide such cases into mild conflicts and dilemmas. In a mild conflict, the agent judges that, *prima facie*, they ought to  $\varphi$  (rather than  $\psi$ ), and that, *prima facie*, they ought to  $\psi$  (rather than  $\varphi$ ). They also judge that they cannot both  $\varphi$  and  $\psi$ . Perhaps she also judges that, overall, she ought to either  $\varphi$  or  $\psi$ . In a dilemma, the agent believes of each option that she ought to choose it. That is, she judges that, overall, she ought to  $\varphi$ .

---

<sup>1</sup>I focus on conflicts about two options.

✉ Nora Heinzelmann  
nora.heinzelmannn@fau.de

<sup>1</sup> Institute for Philosophy, Friedrich Alexander University of Erlangen-Nuremberg, Bismarckstrasse 1, 91054 Erlangen, Germany

<sup>2</sup> Munich Center for Mathematical Philosophy, Ludwig Maximilian University of Munich, Munich 80539 Geschwister-Scholl-Platz 1, Germany

She also judges that, overall, she ought to  $\psi$ . As in a mild conflict, the agent also judges that she cannot both  $\varphi$  and  $\psi$ .

Imagine an agent in a dilemma ends up, say,  $\varphi$ ing. When she  $\varphi$ s, she at the same time judges that she ought to and could have acted differently: she could have  $\psi$ ed. The agent then fulfils the conditions for what is commonly understood as weakness of the will. An agent is weak-willed in failing to  $\varphi$  if she judges, overall, that she ought to  $\varphi$  and that it is possible for her to do so. In a dilemma, the agent is thus weak-willed by definition. But this is puzzling: weakness of will and dilemmas are two different phenomena.

A range of reactions is available to address this puzzle. For one thing, sceptics about weakness of will or dilemmas may see the issue as independent support for their positions. Here, I develop a different response: I argue that plausible accounts of dilemmas and weakness of will can be given that are consistent with one another.

The paper is organised as follows. In section 2, I distinguish mild conflicts from dilemmas. Then I explain (in section 3) why weakness of will requires mild conflict but is inconsistent with dilemmas (section 4). In section 5, I develop and argue for my preferred response to this issue. The concluding section discusses implications.

## 2 Conflicts

### 2.1 Mild Conflicts and Dilemmas

Consider two examples of conflict.

**Simon's Choice** The public swimming pool is about to close but Simon has promised his two small children that he will take each of them in turn for one final ride on the big water slide. He needs to quickly choose whom he takes first: his son or his daughter. On no ground does Simon favour one child over the other. But if he fails to make a choice, the slide will be closed and he cannot keep his promise.

**Sophie's Choice**<sup>2</sup> Upon entering a Nazi concentration camp Sophie is asked to choose which of her two small children, a girl and a boy, she wants to keep. The other will be sent to the gas chamber. On no ground does Sophie favour one child over the other. But if she fails to make a choice, they will both be killed.

Both cases are examples of conflict: each parent has to decide between two children. The parent is not aware of any decisive reasons to choose one option over the other. Yet they are aware that choosing neither option would be wrong, and surely worse than picking one of the two options.

Let ' $\varphi$ ing' be 'picking the son', and ' $\psi$ ing' be 'picking the daughter'. In Simon's case, picking the son amounts to picking him first, and coming back for the daughter afterwards (and *mutatis mutandis* for the other child). In Sophie's case, picking the son amounts to picking the son and abandoning the daughter (and *mutatis mutandis* for the other child).

Each parent sees reasons for  $\varphi$ ing and equally strong reasons for  $\psi$ ing. E.g., it may seem better *in some respect* to pick the son: he is smaller and easier to carry. Yet the parent does not judge that it is better *overall* to pick the son. Conversely, the parent may

<sup>2</sup> Greenspan (1983) takes this case from Styron (1979).

think that it is better *in some (other) respect* to pick the daughter. Perhaps she would get upset if separated from her parent while her brother, immersed in play, would not even notice. Yet the parent does not judge that it is therefore better *overall* to pick the daughter. Neither of the two options seems better than the other, tout court. Each parent also knows that it is impossible to take both children at the same time.

Let us distinguish between two kinds of judgments. *Prima facie* judgments are judgments about what one *prima facie* ought to do, what one has *prima facie* reason to do, or what is in some respect better or best to do. Such judgments do not settle the question of what to do and are not sufficient for action. For instance, a judgment that the son would be easier to carry is not decisive for the parent's choice.

Overall judgments are judgments about what one ought to do, overall, what one has conclusive or most reason to do, or what is in any respect best to do. These judgments take all the agent's relevant evidence into account. Note that an overall judgment is not necessary for action (Bratman 1979, p.161), for instance, in Buridan cases (Zupko 2011, § 7), cases of ambivalence, indifference, or uncertainty.

Simon's and Sophie's cases are thus similar in that the agent makes conflicting *prima facie* judgments. At least one of their *prima facie* judgments favours  $\phi$ ing, at least a second one favours  $\psi$ ing. But the cases also differ. What distinguishes them is this: Simon judges that it is fine to choose one of his children now, as he will have time to come back for the other one afterwards. Perhaps we can ascribe to him the overall judgment that he ought to take either child. Simon's conflict is entirely one of conflicting *prima facie* judgments. He is in what we shall call a mild conflict.

**Mild Conflict** An agent is in a mild conflict iff he judges that, *prima facie*, he ought to  $\phi$ , that, *prima facie*, he ought to  $\psi$ , and that it is impossible to do both.

Sophie's choice is much harder because she wants to save both children and knows that this is impossible. Perhaps we can ascribe to her the judgment that, overall, she ought to save both children, i.e., she ought to both  $\phi$  and  $\psi$ .<sup>3</sup> But given that she knows that this is impossible, we may not wish to say that. What seems clear is her conflict is one of two conflicting overall judgements. On the one hand, Sophie judges that she ought to pick her son. On the other hand, she judges that she ought to pick her daughter. She is in a dilemma.

**Dilemma** An agent is in a dilemma iff she judges that, overall, she ought to  $\phi$ , that, overall, she ought to  $\psi$ , and that it is impossible to do both.

## 2.2 A Clarification

That mild conflicts as I have defined them are conceivable is hardly controversial. In contrast, there is an ongoing debate about whether dilemmas are conceivable.<sup>4</sup> A clarification is therefore in order. This paper understands dilemmas as conflicting

<sup>3</sup> We shall return to this suggestion on pp.4-5.

<sup>4</sup> Those who argue that dilemmas are conceivable include Sinnott-Armstrong 1988, Vallentyne 1989, Hansson 1998, Forrester 1995, Horty 2003, and Bessemans 2011. Conee 1982 and Brink 1994, amongst others, believe that they are not.

mental states, not as conflicting obligations, oughts, duties, norms, or demands. Whether the agent is mistaken or not about her conflicting beliefs, i.e., whether she actually has or lacks those obligations and abilities, is an entirely different matter. Whether the agent is irrational is again a separate issue. The claim we are concerned with here is merely that dilemmas as stated above are conceivable. There are several arguments to support it (see Barcan Marcus 1980; Guttenplan 1980; Forrester 1995; Williams 1965). I shall not argue for it here.

### 3 Weakness of Will and Mild Conflict

Weakness of will<sup>5</sup> requires mild conflict (cf. Davidson 1980 [1970], pp.33–5). But what is weakness of will? Philosophers have struggled to characterise it since ancient times. Although there is no universally accepted definition, it is commonly understood as, or manifested by, action against one's better judgement (Aristotle, *Nicomachean Ethics* 1145a15–35; Mele 2012). More precisely, an agent is weak-willed in  $\phi$ ing iff she judges that she ought to and could do something else instead, e.g., refrain from  $\phi$ ing (Stroud 2014; Davidson 1980 [1970]). Typical examples for weakness of will concern bodily pleasures like food, drink, or sex. For instance, a dieter is weak-willed when he judges that he ought to forego dessert yet gives in to temptation.

Let us thus adopt the following working definition:

**Weakness of Will** An agent is weak-willed iff she judges that, overall, she ought to  $\phi$ , that it is possible for her to  $\phi$ , yet she does not  $\phi$ .

The judgment in this definition is an overall judgment, not merely a prima facie one. The weak-willed person even acts in accordance with a prima facie judgment. The dieter judges that, prima facie, he ought to eat dessert because it is delicious. Still, overall, he judges that despite this pleasure he ought to abstain.

The agent's prima facie judgments that favour the weak-willed action contrast with his prima facie judgments that oppose it. On an Aristotelian view, the agent endorses two conflicting syllogisms which support either the weak-willed action or an alternative that is incompatible with it (Aristotle 1894, *Nicomachean Ethics*, book VII; Aquinas 1912, *Summa theologiae* IIa 156.2). The dieter judges, on the one hand, that pleasure is to be pursued, that eating dessert is a pleasure, and that he therefore ought to eat dessert. On the other hand, he judges that one ought to not taste anything unhealthy, that dessert is unhealthy, and that he therefore ought to not taste dessert.

Davidson (1980 [1970], p.33) famously complained that this approach stops just short of ascribing a straightforward contradiction to the weak-willed agent. On this view, the agent seems to judge that she ought to both do something and not do it. And certainly, we may add, whilst weak-willed agents have some air of self-contradiction about them, they do not merely state a logical contradiction.<sup>6</sup>

<sup>5</sup> I do not distinguish 'weakness of (the) will' and 'akrasia'.

<sup>6</sup> The same objection applies, mutatis mutandis, to the view that an agent in a dilemma judges that she ought to both  $\phi$  and  $\psi$  and that she cannot do both. An agent in a dilemma is not contradicting herself; she faces demands that are each valid on their own but, taken together, cannot all be met.

The Aristotelian view yields this counterintuitive result because it cannot account for mild conflict (Davidson 1980 [1970], pp.33–4). Consider how it would characterise Simon: he reasons, on the one hand, that one ought to act fast because the pond is closing, that taking the son first would be faster, and that he therefore ought to take his son first. On the other hand, Simon reasons that distress in others must be minimised, that taking the daughter first would minimise distress in his children, and therefore that he ought to take the daughter first. Hence, on the Aristotelian view, Simon is making contradicting claims: he ought to take the daughter (and not the son), and he ought to take the son (and not the daughter). This is an inadequate account of mild conflicts (Davidson 1980 [1970], p.24).

Davidson therefore proposes a solution that accounts for mild conflicts as well as weakness of will. He distinguishes two kinds of overall judgments: *all-out judgements* and *all-things-considered judgements*. Both are made in light of all the agent's available evidence and not merely about what one prima facie ought to do. All-things-considered judgments are about what one ought to do relative to one's relevant considerations (p.38). They are thus conditional on the agent's evidence, taking, say, the form 'ϕing is best, all things considered' (p.39). In contrast, all-out judgments are unconditional. Typically, when a person acts intentionally, they act on such a judgment. All-out judgments may, for instance, take the form 'ϕing is best' or 'ϕing is better than ψing.'

On Davidson's account, the weak-willed person judges that, all things considered, she ought to ϕ. But she acts in accordance with an all-out judgment that she ought to ψ, simpliciter, tout court, sans phrase. An agent is thus weak-willed in ϕing for a reason *r* iff she judges that it is possible for her to ψ and ψing is better than ϕing, given some other reason *r*\* that includes *r* and more (Davidson 1980 [1970], p.40).

Despite its popularity, Davidson's view does not allow a weak-willed agent to conclude (all-out) to do one thing but act differently (Bratman 1979; Tenenbaum 1999; Mele 2012; Stroud 2014, p.3.1). This is simply impossible on Davidson's account. Whatever a weak-willed agent does, they do it on a corresponding all-out judgment that it is best to act so.

Yet, pace Davidson, at least sometimes the weak-willed person concludes (all-out) to act prudently but fails. Then the weak-willed agent 'defies' the 'verdict of her practical deliberation' (Buss 1997, p.32–3). This may happen for different reasons. Perhaps they make some (different) mistake in practical reasoning (Bratman 1979). Perhaps they are not persuaded by their reflective understanding although they acknowledge that they ought to be (Tenenbaum 1999, p.902; cf. Tenenbaum 2007, ch.7). Perhaps they 'fail' to decide or intend to do what they believe good or best or right' (Stocker 1979, p.738). Perhaps they do not engage in prior deliberation or form a judgment about what they ought to do, overall; instead, they intend or decide to act in a weak-willed way (Buss 1997; Scanlon 1998, p.47).

All of these accounts distinguish the conclusion of practical reasoning, viz. Davidson's all-out judgment, from a conflicting mental state of a different kind (Bratman 1979; Rorty 1980; Mele 1987; Mele 2003 [1995]; Buss 1997; Mele 2012). This mental state settles the question of what to do. It may be a judgment – perhaps of the form 'I shall ϕ' – that differs in some crucial respect from Davidson's all-out judgment (Bratman 1979, p.163). Or it might be a decision or intention (Mele 2003 [1995], p.71). In all cases, the weak-willed agent acts in accordance with such a mental state and against her all-out judgment.

At the same time, these accounts allow for mild conflict because they allow for conflicting *prima facie* judgments which are not sufficient for action. Still, they do not ascribe a straightforward contradiction to the weak-willed agent. There is an air of self-contradiction about him because of his conflicting mental states, like ‘I ought to not eat dessert’ and ‘I shall eat dessert’. However, because these conflicting states are of different kinds, they are not logically contradictory. One of them is an overall judgment, the latter a different kind of judgment, an intention, or a decision.

## 4 Weakness of Will and Dilemmas

The goal of this paper is to decide whether a coherent account of dilemmas is consistent with a coherent account of weakness of will. More precisely, it aims to resolve a puzzle that we can in Davidsonian spirit couch as an inconsistent triad:

- (1) An agent is in a dilemma iff she judges that, overall, she ought to  $\varphi$ , that, overall, she ought to  $\psi$ , and that it is impossible to do both.
- (2) An agent is weak-willed if she judges that, overall, she ought to  $\varphi$ , that it is possible for her to  $\varphi$ , yet she does not  $\varphi$ .
- (3) In a dilemma, it is not the case that agents are necessarily weak-willed.

(1) states the account of dilemmas discussed in the first part of this paper. (2) follows from the common definition of weakness of the will discussed in the previous section (the conditional is sufficient). (3) states that dilemmas and weakness of will are different phenomena, a claim we shall discuss in section 4.4.

The three claims are jointly inconsistent. In what follows I first discuss four unsuccessful attempts to resolve the puzzle before presenting a solution.

### 4.1 Action against Comparative Judgments

This response rejects (2) and proposes a different account of weakness of will. This modified account understands the overall judgment that the weak-willed agent acts against in such a way that it differs meaningfully from the overall judgments in dilemmas. There are at least two ways to spell this out, which I shall consider separately. According to the first proposal, the agent’s overall judgment that she ought to  $\varphi$  is to be understood as a judgement that she ought to  $\varphi$  rather than  $\psi$ , where  $\psi$ ing is any salient alternative to  $\varphi$ ing. As agents can only consider a limited range of possible ways to act, a salient alternative is, roughly, any relevant possible action not identical to  $\varphi$ ing, such as  $\neg\varphi$ ing. This is, in fact, Davidson (1980 [1970])’s proposal.

The response continues that agents in dilemmas do not genuinely make overall judgements of this comparative kind. E.g., Sophie does not judge that, overall, she ought to save her daughter rather than her son (and vice versa). Indeed, such a combination of overall comparative judgments may be psychologically impossible. Therefore, Sophie does not fulfil a necessary requirement for weakness of will. The puzzle does not arise.

This reply is viable but proposes an account of weakness of the will that has been superseded in the literature for independent reasons. That it might solve our present problem

would be a reason to resurrect it but it seems to me that it is greatly outweighed by at least three reasons against doing so.

First, as mentioned above, critics have objected that the view does not allow for cases of weakness of will where the agent does not (merely) act against such a comparative overall judgment but rather against a different kind of mental state. However, a plausible account of weakness of will should allow for such cases. For instance, it ought to allow for cases where an agent concludes (all-out) to do one thing but still does something else (Bratman 1979; Stocker 1979; Mele 1987; Buss 1997; Scanlon 1998; Tenenbaum 1999; Mele 2012; Stroud 2014). A dieter who judges all-out that he shall decline dessert may well be weak-willed when he then succumbs to temptation after all.

At this point, one might suggest that we should use different accounts with different labels for similar but still different phenomena.<sup>7</sup> For instance, let us use “*weakness of will 1*” to refer to action against an all-out judgement and “*weakness of will 2*” for action against a comparative judgement. The puzzle is not a problem for “*weakness of will 2*”. But it is still a problem for “*weakness of will 1*”. Hence, the puzzle merely shifts in scope, it does not go away. *Weakness of will 2* is implausible but avoids the puzzle. *Weakness of will 1* is more plausible than *weakness of will 2* but faces the puzzle. Neither of the two accounts is plausible on all counts. I suggest that we should use yet another account that avoids both the puzzle and the issues that *weakness of will 1* faces, and is independently plausible. I shall present this account in section 5.

Second, the definition, as stated, is technically incorrect (Mele 2012, p.5). Imagine the dieter judges that, overall, abstaining is better than indulging but even better than abstaining is indulging and going for a run after dinner. Now, if the agent eats dessert, he acts against his better judgment that he ought to abstain. Provided he judges that he is able to do otherwise, he is weak-willed by the definition under consideration. Still, if the dieter goes for a run after dinner, he is not weak-willed if he eats dessert.

Lastly, imagine the dieter is a satisficer who thinks that skipping dinner entirely would be the best thing to do but at the very least he ought to skip dessert. I.e., he judges that, overall, foregoing dinner is better than foregoing dessert. Now imagine the dieter has dinner but not dessert although he judges that it would be possible for him to skip dinner entirely. On the definition under consideration, he is weak-willed: he fails to act on an overall comparative judgment that a salient alternative is better. But this seems incorrect. Clearly, the satisficer does what he judges to be good enough, even though there is a better option available to him. Again, the comparative account is inadequate.

According to the second proposal, the agent’s overall judgement is to be understood as a judgement that she ought to  $\varphi$  but that it is not the case that she ought to  $\psi$ . This yields the following account of weak-willed action<sup>8</sup>:

- i.  $S$  judges that she ought to  $\varphi$  but that it is not the case that she ought to  $\psi$ ,
- ii.  $S$  judges that she can  $\varphi$  and that she can  $\psi$  (but not both), and
- iii.  $S$   $\psi$ s (rather than  $\varphi$ s).

This account avoids the puzzle. Consider Sophie again: she judges that she ought to save her son. She does not judge that it is not the case that she also

<sup>7</sup> I thank an anonymous reviewer for helping me to clarify this point.

<sup>8</sup> I thank an anonymous reviewer for suggesting this account.



ought to save her daughter. So, she does not fulfil condition (i). She is not weak-willed. The puzzle does not arise.

However, the account has other problems. For one thing, it is too broad. Imagine *S* is on a diet. There are several options for dessert: fruit salad, creamy chocolate cake, and sundae. *S* judges that she can eat fruit, cake, nothing, etc. *S* judges that she ought to choose fruit rather than cake or sundae. *S* also judges that it would be even better to skip dessert altogether but this would be supererogatory: *S* judges that it is not the case that she ought to skip dessert. Imagine that *S* skips dessert altogether. Is *S* weak-willed? It seems to me that *S* is not weak-willed. After all, *S* is doing something that she judges to be supererogatory.

According to the account given above, *S* fulfils condition (i): *S* judges that she ought to eat fruit but that it is not the case that she ought to skip dessert. *S* also judges that she can eat fruit and that she can skip dessert (but not both). So, condition (ii) is fulfilled as well. *S* skips dessert instead of eating fruit, fulfilling condition (iii). *S* is weak-willed on the account under discussion.

A related class of counterexamples concern cases where an agent's  $\phi$ ing depends on success conditions outside of her control. Imagine *S* judges that she ought to win the race but that it is not the case that she ought to come second. *S* also judges that she can win the race, and that she can come second (but not both). Imagine that *S* tries to win the race, and comes second by a very short margin. *S* thus fulfils all conditions (i)-(iii) for weakness of the will. Yet *S* is not weak-willed. The account is too broad.

Furthermore, the account also seems too narrow. Imagine another dieter *R* who judges that she ought to eat fruit salad but not cake. Imagine also that *R* does not judge that it is not the case that she ought to eat cake. Perhaps *R* does not make a judgement about the cake at all, or perhaps *R* does not make a more specific judgement about what she ought or ought to not do regarding the cake. Then *R* does not fulfil condition (i). Imagine that *R* judges that she can have cake or fruit (but not both), thus fulfilling condition (ii), and that she ends up eating the cake, fulfilling condition (iii). According to the account, she is not weak-willed. But plausibly, she may be: she fails to eat the fruit and thereby violates her diet.

## 4.2 Dilemmas Entail Compulsion

Another response rejects (1) and argues that dilemmas entail compulsion, that weakness of will is different from compulsion, and that, therefore, an agent in a dilemma need not be weak-willed. Here, I shall accept the second premise of this argument but reject the first. I acknowledge that, in a dilemma, the agent is forced by external circumstances to fall short of doing what they would like to do. But this is not compulsion. For instance, because Sophie wants to save both children and this is impossible, she is compelled to fail by her own lights. Still, Sophie is free to choose between her children, she can pick the girl or the boy. She is constrained only in the sense that her choice set is restricted. That is, picking both children is not an option she can select. In contrast, in cases of compulsion like obsessive-compulsive disorder (OCD), the patient is not free to decide between, say, washing her hands or not. Her choice set is constrained to such an extent that she has but one option, namely to continue washing her hands.

Note that the number of options is not crucial: imagine an OCD patient suffers from two compulsions, to wash her hands, and to check whether the oven is on. Imagine she



alternates between the two, running from the bathroom to the kitchen and back again. In a way, this patient has two choice options, just like Sophie. Still, her case is substantially different: when she stops washing her hands and rushes to the oven, she is compelled to do so. She is not making a choice. Perhaps she is not even performing an action at all because she does not wash her hands or check the oven *intentionally*. In contrast, Sophie does make an autonomous choice between her children. Her bodily movements are under her own control. She does not act from compulsion.

The following variation illustrates this:

**Obsessed Sophie** Sophie suffers from a strange obsession. When asked a question about her children, she automatically and exclusively talks about her son. Even in situations where she knows that she ought to bring up her daughter – e.g., when asked whether she has girls – she only mentions her son. So, the concentration camp officers request her to pick one of her children, she mechanically chooses her son.

Obsessed Sophie is picking her son out of compulsion. She is not weak-willed: it is not possible for her to do otherwise. Still, her case is importantly different from non-obsessed Sophie's. Even if non-obsessed Sophie picks her son, she does not do so because she suffers from a mental disorder. She is compelled to abandon one of her children, yet she is not compelled to move or talk in any particular way. Hence, non-obsessed Sophie's dilemma and dilemmas more generally do not entail compulsion.

### 4.3 Acting Unintentionally

This response also rejects (1) and claims that, in a dilemma, the agent acts unintentionally. As acting intentionally is a requirement for weakness of will, an agent in a dilemma is not weak-willed.

In response, I reject the first premise of this argument. If it was correct then Sophie would pick her son unintentionally. This might happen because an action can be intentional under some description but not under another (Anscombe 1957). E.g., Sophie, not speaking the staff's language, might think she was asked which child is younger when instead they asked whom she wants to keep. To this, she might reply 'my son'. Let us use ' $d_1$ ' to designate 'picking the son' and ' $d_2$ ' to designate 'naming the younger child'. Sophie's utterance is intentional under description  $d_2$  but not under  $d_1$ . She is unintentionally picking her son. But she does not act against her judgment that, overall, she ought to pick her daughter because this judgment features  $d_1$ , not  $d_2$ . She is thus not weak-willed.

However, this is not the puzzling case that concerns us in this paper. Our case is one where the agent acts intentionally under some description  $d$  but has conflicting overall judgments featuring the *same* description  $d$ . But in this case, the puzzle does arise. Considering a different case that is not puzzling does not solve the problem.

### 4.4 Biting the Bullet

This suggestion rejects (3) and accepts that agents in dilemmas are necessarily weak-willed. After all, whatever the agent does, she judges that she ought to not do it – otherwise, the situation would not be a dilemma. But precisely not

doing what you judge you ought to do is being weak-willed, provided you fulfil all other requirements.

This reply faces two problems. First, the move is intuitively implausible. Dilemmas and weakness of will are different phenomena. A variation of Simon's and Sophie's case illustrates this:

**Drink** The agent happens to have a bottle of gin in their bag, bought for an upcoming party. About pick one of their children, they consider quickly taking a sip of the gin, which it would make them more relaxed. At the same time, they judge that they absolutely must not lose time, and drinking would cost precious moments. But the temptation to postpone the decision for just a few seconds and to escape to a moment's pleasure is too great: they quickly take a sip.

In drinking alcohol instead of quickly picking a child, the agent is plausibly weak-willed. This is because they act against their judgment that, overall, they ought to not drink. Drinking alcohol is worse than picking either child.

Why is drinking a genuinely weak-willed action but, say, taking the son is not? In both cases, the parent acts against their better judgment that they ought to do something else, namely, take the daughter and abstain. They judge that, in some respect, it would be better to take the daughter. Similarly, they judge that, in some respect, it would be better to drink: it would make them more relaxed. In both cases, they are weak-willed by definition. But in Drink, this verdict is plausible. In the case of picking one of the children, it is not. Therefore, weakness of will is crucially different from a dilemma. An agent may be in a dilemma, and they may *in addition* be weak-willed, as in Drink, yet they are not necessarily weak-willed in any dilemma.

Second, biting the bullet not only amounts to granting that an agent in a dilemma is necessarily weak-willed. It also implies that she is necessarily strong-willed at the same time. The strong-willed agent makes the same judgments as the weak-willed one, yet manages to overcome temptation and to act as they think they ought to. For instance, whatever Sophie does, she will act in accordance with one of her judgments. So, she is strong-willed by definition. Yet, at the same time, she is weak-willed, too.

If being strong-willed and being weak-willed are opposites, it is implausible that an agent could be both at the same time, in doing one and the same thing, and it is even more implausible that this should be necessarily so. Note that we are not concerned with a case where an agent is weak-willed at one time or with regard to one action and strong-willed at another time or with regard to another action. For instance, someone might fail to adhere to a dieting rule but fail to turn down the offer of a cigarette. That such cases are plausible is rarely contested, yet they are not at issue here. The cases discussed here are also not to be conflated with examples in which an agent performs two actions, one in line with and one against his better judgement (Aristotle, *Nicomachean Ethics* VII.9; Mele 2003 [1995]; Holton 1999). E.g., a boy yielding to peer pressure might decide to break into a house against his better judgement but then 'chicken out' at the last moment (Mele 2003 [1995], pp.61–63). Given that he decides to act against his better judgement, he might be called weak-willed (if deciding is an action). Given that he does not end up breaking into the house, he might be called strong-willed. But note that the deciding and the 'chickening-out' are two different actions. Here, we are concerned with just one action and two conflicting better judgements.

However, there is one important lesson here: weakness of will can be relative to a time and action. By extension, perhaps it can be relative to a judgment about what one

ought to do? If so, one might argue that an agent can be weak-willed with regard to one such judgement and strong-willed with regard to another one. For instance, in picking her son, Sophie would be weak-willed regarding her judgment that she ought to pick her daughter but strong-willed regarding her judgment that she ought to pick her son. Thus, an agent can no longer be said to be weak-willed or strong-willed simpliciter but only regarding one of her judgements.

On this view, weakness and strength of will apply to a very wide range of cases. For instance, in doing philosophy, I am weak-willed with regard to my judgement that I ought to fight world poverty but strong-willed with regard to the judgement that I ought to contribute to humanity's quest for truth. Whilst this does not seem plainly absurd, stating that someone is weak-willed or strong-willed would then appear as an innocuous ascription. It might be legitimately asked whether, understood in this way, the concepts would retain much of the meaning we would desire them to bear. In addition, we would need to support this view independently of considerations concerning the problem discussed in this paper. I think there is a better response.

## 5 Solution

My preferred response to the puzzle rejects (2) and favours a different account of weakness of will. On this account weak-willed action is not action against an overall judgement but against a different mental state. This account is consistent with the account of dilemmas.

To begin, let us distinguish the mental states that are involved in dilemmas from those that are involved in weakness of will. In a dilemma, the agent's judgments about what she ought to do, overall, are neither necessary nor sufficient for action. For instance, Sophie judges that she ought to save her daughter and that she ought to save her son. What she does is not settled by these judgments. In contrast, in a case of weakness of will, the agent acts against a different kind of mental state, which *does* settle the question 'What shall I do?' An agent is weak-willed in failing to act accordingly. This kind of mental states must meet at least the following three criteria:

- (a) It is not sufficient for action. Otherwise, the weak-willed agent could not act against it.
- (b) The weak-willed agent must not act on a token of this kind of mental state that conflicts with another token of the same kind. Otherwise, the puzzle could arise again: if a case were coherent where an agent had two conflicting mental states of this kind and she acted against one of them, she would be weak-willed by definition.
- (c) Having a mental state of this kind must allow for being in a mild conflict. As we have seen above, in a case of weakness of will there may be two conflicting *prima facie* judgments, such as that it is better in one respect to perform the weak-willed action (say, indulge in dessert because it is pleasant), and that it is better in some other respect to not perform this action (say, to abstain for better health). This mild

conflict may easily arise even if there was some further kind of mental state, such as an intention or decision to, say, indulge.

In our quest to find a kind of mental states that has these three properties, let us first consider certain judgments. Most judgments are unsuitable because they do not fulfil (b): the weak-willed agent can act in accordance with a token of this kind of judgment that conflicts with another token of the same kind. This is true for all *prima facie* and at least some overall judgments. However, some judgments may fulfil (b). For instance, it seems that an agent cannot judge that, all things considered, he ought to  $\varphi$  and that, all things considered, he also ought to not  $\varphi$ . One might thus be tempted to suggest that an agent is weak-willed if she acts against a judgment about what to do, all things considered. However, such an account is all-too-familiar (Davidson 1980 [1970]) and has been criticised for failing to account for cases of weakness of will where an agent reasons correctly but still acts against her conclusion. Hence, we might wish to bracket all-things-considered judgments as candidates for the kind of mental states we are looking for. Davidson's all-out judgments are not promising either because they do not fulfil (a). On Davidson's account, it is not possible for the weak-willed agent to act against an all-out judgment.

Similar considerations apply to some of the judgments that feature in attempts of Davidson's critics to explain how a weak-willed agent may act against a conclusion of practical deliberation (mentioned above). E.g., Bratman's 'practical judgments' of the form 'I shall  $\varphi$ ' settle the question of what to do. Although they are products of practical reasoning, they differ essentially from overall judgments, which are evaluative (Bratman 1979, p.163). These judgments do not meet (a): the weak-willed agent cannot act against them.

Let us turn to mental states that are not judgments. Authors have suggested decisions and intentions. They too are supposed to settle the practical question 'What shall I do?' (Mele 2003 [1995], p.72) and commit the agent to a certain course of action. They might not need prior deliberation or an overall judgment about what to do (Scanlon 1998, p.47) but they are often formed through such an inferential process (Mele 2003 [1995], pp.71–2).

These mental states seem to fulfil all three criteria. First, they are not sufficient for action: an agent might decide or intend to do something, like declining dessert, and yet fail to act accordingly. Second, when the weak-willed agent acts in accordance with an intention or decision, they do not at the same time endorse a conflicting intention or decision. Indeed, it seems impossible to intend or decide to do something and at the same time intend or decide to do something that is incompatible with it. It is entirely possible that, say, I have promised to go to London on the 14th and to Sydney on Wednesday and accordingly intend to do both, without realising that the 14th is Wednesday. Yet I do not at the same time endorse two conflicting intentions because my intentions do not conflict, from my point of view. Third, it is possible to intend or decide to  $\varphi$  whilst being in a mild conflict. For instance, the parent might decide and intend to take the son without thereby suspending the judgments that they *prima facie* ought to take their daughter because this would minimise distress, etc. Similarly, the dieter, even whilst being firmly decided on declining dessert, judges that, *prima facie*, indulging would be a pleasure but not doing so would be healthy. Hence, intending or deciding to  $\varphi$  is coherent with having conflicting *prima facie* judgments about whether to  $\varphi$  or not.

This solves the puzzle. In a dilemma the agent judges that, overall, she ought to  $\varphi$ , she judges that, overall, she ought to  $\psi$ , and she judges that it is impossible for her to both  $\varphi$  and

$\psi$ . An agent is weak-willed iff she intends or decides to  $\varphi$ , judges that it is possible for her to  $\varphi$ , yet does not  $\varphi$ . An agent in a dilemma is thus not necessarily weak-willed because she does not act against an overall judgment that she ought to  $\varphi$  (or  $\psi$ ).

This proposal implies that weak-willed action is to be understood as action against the agent's intention or decision. Although this view has so far been that of a minority in the literature on weakness of will, it has recently gained support. E.g., Mele (1987, 2003 [1995], 2012) characterises some cases of weakness of will as failure to act against an intention or a decision. When a boy decides and intends to break into a house against his better judgment but then “chickens out”, this chickening-out may be weak-willed because it thwarts the boy's decision and intention (Mele 2003 [1995], p.71). In these cases, ‘a practical commitment is thwarted by non-compelling competing motivation’, just as in ‘orthodox’ cases of action against one's better judgment (Mele 2003 [1995], p.74). In the same vein, Holton (1999, 2009) describes weakness of will as over-ready suspension of one's previously formed resolution in the face of countervailing temptation. A resolution is a pair of intentions: an intention to  $\varphi$  and a meta-intention to persist in that. Not all revisions of one's intention are weak-willed. For one thing, when the agent gains relevant new evidence, she may be permitted to revise the resolution (Holton 2009, p.75).

A difficult question is to what extent unexpected circumstances count as relevant new evidence. Not just any unexpected differences will do. For instance, if the dieter anticipated that dessert would be a heavy chocolate brownie and decided to decline it, he may still be weak-willed when he indulges in the equally heavy caramel cream that is served instead. Similar considerations apply when dessert is simply *more* tempting than expected. Roughly, it seems, that new evidence is relevant only if it concerns the reasons the agent's decision and intention were based on. E.g., in the dieter's case, new evidence pertaining to health aspects of dessert is relevant but new evidence on its tastiness or other features unrelated to health is not.

## 6 Concluding Remarks

Weak-willed action is best understood as action against one's decision or intention. To the arguments in favour of this view this paper has added another one: this account does not fall prey to the threat that agents are necessarily weak-willed in a dilemma, i.e., when they are conflicted about what they ought to do, overall.

Where does this leave action against one's better judgment? Note that, in most cases, this action will not only be against one's better judgment but in addition against an intention conforming with that judgment. Hence, the agent will be weak-willed. But not only are there cases where intention and better judgment diverge. We also wish to say something about an agent's mere failure to intend to do what she judges she ought to do. For instance, imagine the dieter judges that, overall, he ought not eat dessert but then intends to have it anyways. Regardless of what he ends up doing, it seems his intention is flawed in the first place.

We need not call this flaw ‘weakness of the will’ (Mele 2003 [1995], p.63). Moreover, we can still say that, whatever label we may use for this phenomenon, it is irrational to not intend to do what you judge you ought to do (Broome 2013; Fink 2013, special issue on Enkrasia). It follows that, if an agent is in a dilemma, she seems to be necessarily irrational in this way (Kiesewetter 2018). One might wish to react to this challenge in a way similar to one of the reactions to the puzzle presented here. For instance, one might wish to reject

dilemmas, the claim that failing to intend to do what you judge you ought to do is irrational, or the claim that, in a dilemma, the agent is not necessarily weak-willed. But this will be a different discussion in a somewhat different field. Here, we cannot pursue it any further.

**Acknowledgements** For comments on previous versions of this paper I thank Arif Ahmed, Richard Holton, and Guy Kahane. For helpful discussion, I thank John Broome, Michael Cuffaro, Alex Greenberg, Erasmus Mayr, Regina Rini, and audiences in Cambridge, Karlsruhe, Münster, Munich, Osnabrück, Oxford, and Tübingen. For financial support I thank the Arts and Humanities Research Council UK (grant no 04386), the University of Cambridge, and the German Academic Exchange Service (DAAD).

**Funding Information** Open Access funding provided by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anscombe, E. (1957). *Intention*. Cambridge (MA): Harvard University Press.
- Aquinas, T. (1912 [1265–73]). *Summa theologiae*. Ed. fathers of the English Dominican Province, London: Burns Oates and Washbourne.
- Aristotle (1894 [ca. 350 BC]). *Nicomachean Ethics*. ed. I. Bywater, Oxford: Oxford University Press.
- Barcan Marcus, R. (1980). Moral dilemmas and consistency. *Journal of Philosophy*, 77(3), 121–136.
- Bessemans, C. (2011). Moral conflicts and moral awareness. *Philosophy*, 86, 563–587.
- Bratman, M. (1979). Practical reasoning and weakness of the will. *Noûs*, 13(2), 153–171.
- Brink, D. (1994). Moral conflict and its structure. *Philosophical Review*, 103(2), 215–247.
- Broome, J. (2013). *Rationality through reasoning*. Oxford: Blackwell.
- Buss, S. (1997). Weakness of will. *Pacific Philosophical Quarterly*, 78(1), 13–44.
- Conce, E. (1982). Against moral dilemmas. *Philosophical Review*, 91(1), 87–97.
- Davidson, D. (1980 [1970]). *How is weakness of the will possible?* In *his Essays on actions and events* (pp. 21–42). Oxford: Clarendon.
- Fink, J. (2013). Editorial. *Organon F*, 20(4), 422–424.
- Forrester, J. (1995). Conflicts of obligation. *American Philosophical Quarterly*, 32(1), 31–44.
- Greenspan, P. (1983). Moral dilemmas and guilt. *Philosophical Studies*, 43(1), 117–125.
- Guttenplan, S. (1980). Moral realism and moral dilemmas. *Proceedings of the Aristotelian Society*, 80, 61–80.
- Hansson, S. (1998). Should we avoid moral dilemmas? *The Journal of Value Inquiry*, 32, 407–416.
- Holton, R. (1999). Intention and weakness of will. *Journal of Philosophy*, 96(5), 241–262.
- Holton, R. (2009). *Willing, wanting, waiting*. Oxford: Clarendon.
- Horty, J. (2003). Reasoning with moral conflicts. *Noûs*, 37(4), 557–605.
- Kiesewetter, B. (2018). Contrary-to-duty scenarios, deontic dilemmas, and transmission principles. *Ethics*, 129(1), 98–115.
- Mele, A. (1987). *Irrationality*. New York: Oxford University Press.
- Mele, A. (2003 [1995]). *Autonomous agents: From self-control to autonomy*. New York: Oxford University Press.
- Mele, A. (2012). *Backsliding*. Oxford: Oxford University Press.
- Rorty, A. (1980). Where does the akratic break take place? *Australasian Journal of Philosophy*, 58(4), 333–346.
- Scanlon, T. (1998). *What we owe to each other*. Cambridge (MA): Belknap Press.
- Sinnott-Armstrong, W. (1988). *Moral dilemmas*. Oxford: Blackwell.

- Stocker, M. (1979). Desiring the bad: An essay in moral psychology. *Journal of Philosophy*, 76(12), 738–753.
- Stroud, S. (2014). 'Weakness of will'. Stanford Encyclopedia of Philosophy. Ed. E. Zalta. Spring 2014.
- Styron, W. (1979). *Sophie's choice*. New York: Random House.
- Tenenbaum, S. (1999). The judgment of a weak will. *Philosophical and Phenomenological Research*, 875–911.
- Tenenbaum, S. (2007). 'Appearances of the good: An essay on the nature of practical reason'.
- Vallentyne, P. (1989). Two types of moral dilemmas. *Erkenntnis*, 30(3), 301–318.
- Williams, B. (1965). 'Ethical consistency'. Proceedings of the Aristotelian Society. *Supplementary*, 39, 103–138.
- Zupko, J. (2011). 'John Buridan'. Stanford Encyclopedia of Philosophy. Ed. E. Zalta. Fall 2011.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.