

Journal of Social Sciences and Philosophy

1995(September); Vol.7, No.2, pp.277-311

© Sun Yat-Sen Institute for
Social Sciences and Philosophy,
Academia Sinica

Freedom and *Akrasia*

Jih-Ching Ho*

* Assistant Research Fellow, Institute of European and American Studies, Academia Sinica

(Received: February 9, 1995; Accepted: April 12, 1995)

Abstract

There are free, weak-willed actions. It is an apparent fact that we sometimes act freely against our best judgments—we sometimes perform actions in our own free will, while judging that another possible action would have been, overall, a better one. However, a widely held skeptical view has it that one performs an action against one's own best judgment only if one does so unfreely; in other words, there can be no free weak-willed action. In this paper, three main arguments for this form of skepticism are examined. In order to defend the possibility of free weak-willed actions, I explain why all these arguments fail.

Outline

- I. The No Ability Argument
- II. Plato's Skepticism
- III. Pugmire's Skepticism
- IV. The Unexercisability Argument
- V. Watson's Skepticism (I)
- VI. Watson's Skepticism (II)
- VII. Concluding Remarks

I. The No Ability Argument

An agent performs a weak-willed action (an *akratic* action) under the following conditions: an agent, S, believes that there are two options for action, A and B, open to him and he is free to perform one but not both; S judges that, all things considered, it would be better to do A than to do B; but nevertheless S intentionally does B. In brief, in a case of weakness of will (akrasia) an agent acts intentionally, freely, and knowingly against his own best judgment.

This phenomenon seems familiar enough, for we know, or at least our pre-theoretic intuition grants, that sometimes our actions fit the above description of weakness of will. Since Plato, however, this intuition has been challenged by various philosophical theories about human rationality and action. According to these theories, our concepts of freedom and intentionality of human action exclude the possibility of weakness of will. Apparent weak-willed actions, on this account, are either unintentional, unfree, or simply incomprehensible.

The central issue about weakness of will is not only to show that it is possible that weakness of will occurs, but also to give a cognitively revealing explanation of its occurrence. An approach that I think would serve these purposes particularly well is the *Motivational Strength Approach*.¹ The underlying assumption of this approach is that it is possible that one's evaluation and motivation do not coincide. Under the framework of the Motivational Strength Approach, the basic structure of weakness of will can be described as follows: an agent, S, has conflicting reasons, R1 and R2; S evaluates R1 as weightier than R2; but the motivational strength of R2 is, at the time of action, greater than that of R1, and accordingly S acts on R2. The idea behind this analysis is that an agent's evaluation of a reason (or desire) can be disproportionate to the strength of the reason (or desire) in motivating the agent to act. That

is, a reason may be evaluated higher than its alternative but have less motivational strength, or vice versa. When such a disproportion occurs, there is a split between the agent's evaluation and his motivation, and a weak-willed action is done as a result.

The adequacy of the Motivational Strength Approach depends on many issues, such as whether a split between one's evaluation and motivation is essential for weakness of will, whether there are acceptable explanations for the occurrence of a split of this sort, and whether the approach allows free weak-willed actions. In this paper I consider only the last issue, the issue of freedom and *akrasia*. A widely held challenge to this motivational strength approach claims that under the analysis of the approach, weak-willed actions are indistinguishable from compulsive actions. Given that compulsive actions are done unfreely, it would follow that there can be no free weak-willed action. This type of skeptical view can be traced back to Plato, and has been developed by contemporary philosophers. In this paper, I shall examine their grounds for holding this view and, in order to show that there can be free weak-willed action, explain why their arguments should be rejected.

Plato assimilates weak-willed actions to compulsive actions, because he maintains that a weak agent has *no ability* to resist the desire to perform the less preferred action. The contemporary followers of Plato's theory such as D. Pugmire and G. Watson apparently adopt this viewpoint and both offer their own versions of no ability arguments (Pugmire, 1982; Watson, 1977). The central claim they all share is that a person who succumbs to a temptation and thus knowingly acts against his best judgment has no ability to resist the temptation. If their view is right, then there is a significant similarity between *akratic* actions and compulsive actions, namely, that they are both done under irresistible desires. It follows from this that *akratic* actions, like compulsive actions, are unfree. I will formulate what I take to be the central argument for the

skeptics as follows:

The No Ability Argument

- (P1) If an agent performs a weak-willed action (i.e., he succumbs to temptation and acts against his best judgment), then he acts on a desire which he has no ability to resist.
- (P2) If an agent acts on a desire which he has no ability to resist, then his action is done under an irresistible desire.
- (P3) Therefore, all akratic actions are done under irresistible desires.
- (P4) If an action is free, then it is not done under an irresistible desire.
- (C) Therefore, all akratic actions are unfree.

The argument is valid. The most questionable premise is (P1).² In the following, I will examine some arguments which attempt to vindicate this premise.

II. Plato's Skepticism

In *Republic*, Plato advocates a tripartite conception of the soul according to which human soul contains three parts: the Intellect, the Appetite, and the Emotion. Each part has its own functions, desires, and guiding principles. In *Phaedrus*, Plato claims the following (237d):

We must observe that in each one of us there are two ruling and leading principles, which we follow withersoever they lead; one is the innate desire for pleasure, the other an acquired opinion which strives for the best.

In Plato's view, these two principles are "principle of pleasure" and "principle of the good," and they govern different parts of the soul. The Intellect is governed by the "principle of the good" and thus desires from the Intellect are "desires for the good"; on the other hand, the leading principle in the Appetite or Emotion are "principle of pleasure"

and their desires are “desires for pleasure”(439a-441c).

Since different parts of the soul have their own desires and guiding principles, they may conflict with each other. Plato maintains that the Intellect is supposed to be the guiding part of the soul: if a person is in a normal (or “just,” in Plato’s word) condition, the Intellect governs the Appetite and Emotion. However, Plato professes that there can be a “strife” between different parts of the soul:

These two [the principle of the good and the principle of pleasure] sometimes agree within us and are sometimes in strife; and sometimes one, and sometimes the other has the greater power (238a).

Talking about *power*, we should note that the distinction between the evaluative weight of a reason for action and its motivational strength, on which the Motivational Strength Approach is based, has a counterpart in Plato’s theory. In many places, Plato mentions not only the “assessment” of a desire, but also its “power.” For Plato, an agent’s assessment of a desire is made on the basis of the agent’s principle of the good, and the power of a desire depends on how the desire motivates an agent to act. Consider the case in which an agent’s desire for the good conflicts with a desire for pleasure. There is no doubt that the assessment of the desire for the good is always higher than that of the desire for pleasure, since the former, not the latter, arises from the Intellect, the part of the soul ruled by the principle of the good. The question concerning us is, do desires for the good always have greater power than desires for pleasure? The answer is no. As the above quotation shows, the principle of the good does not always have the greater power.

Plato’s theory of weakness of will, in sum, is that human beings are constantly in the conflict between reasons and desires (or, in his terminology, the strife between desires for the good and desires for pleasure),

and when a person is not in a just condition, it is possible that his desire for pleasure overcomes his desires for the good by having greater causal power. In other words, weakness of will is characterized as the result of the usurpation of desires from the Appetite or the Emotion over desires from the Intellect, in short, the usurpation of desires over reasons.³

So much for Plato's theory of akrasia. We may now consider the concept of freedom. Plato holds that if an agent acts against his own best judgment as a result of a divergence between the agent's assessment and the causal power of his desires, then the agent is not free in performing the action. As Santas points out (1966), Plato apparently embraces the following view:

When a man acts contrary to his knowledge or belief of what is best (for him), and the true explanation of his action is in terms of the strength of his conflicting desires, . . . then the agent acted under psychological compulsion or was psychologically unable to refrain from doing what he did (p. 31).

Is there any reason to suppose that the explanation of weakness of will in terms of divergences between an agent's assessment of a desire and the motivational power of the desire necessarily analyzes weak-willed actions as compulsive actions? In my view, there may be two accounts why on Plato's theory weak-willed actions are construed as unfree.

The first account is that, for Plato, the Intellect is the only part of the soul governed by the principle of the good; so, when an agent performs a weak-willed action, the agent acts on a desire springing from the part of soul which is independent from the agent's principle of the good. That is to say, the agent acts on a desire which has nothing to do with what he thinks is good. From this perspective, weak-willed actions share an important feature with compulsive actions.⁴

An immediate problem is that it is inappropriate to describe weakness of will as a result of a war between reason and desire. A commonly accepted analysis of practical reason, generally attributed to Hume, is the so-called “belief-desire model” of practical reason. One of the simplest forms of the belief-desire analysis, which Bernard Williams describes as “the sub-Humean model,” is the following analysis (Williams, 1979: 101-2):

A person S has a reason to perform an action A if and only if S has some desire the satisfaction of which S believes will be served by his doing A.

That is to say, an agent’s reason for doing A is the agent’s desire for some states of affairs which he believes can be brought about through his action, and his belief that his doing A will contribute to attain that states of affairs. In brief, a reason for action is basically a consideration consisting of a desire and a connecting belief. So construed, a desire, usurping or not, may very well constitute a reason for action. It follows that what is essential for weakness of will is not that one acts on a desire contrary to one’s reason, but that one acts on a reason which is less supported by one’s evaluation. It is unnecessary and inappropriate to distinguish different species of desires.

The second account, as I shall explain below, merits more attention. According to Plato the Intellect is the highest faculty of reason and should dominate other faculties of the soul; thus, if the Intellect cannot prevent the usurpation of the Appetite or the Emotion, then nothing can. On this account, weak-willed agents have no ability to prevent themselves from performing their actions.⁵

A critical problem with this platonic account is that it assumes a false principle. Plato’s main contention leading to skepticism is the following principle which may be called *Plato’s principle*: if one fails to

resist a usurping desire, then one is unable to resist the desire. Without further elaboration, Plato's principle is clearly false.⁶ Generally speaking, it is not the case that when one fails to do something, one is unable to do it. In any event, the fact that an agent fails to do something does not rule out the possibility that the agent has the ability to do it, since this fact is consistent with the hypothesis that *the agent does not exercise the ability*.

Plato's principle, however, can be made plausible in the following two ways, both relevant to the hypothesis of an unexercised ability. (a) The principle can be modified by adding a condition to the antecedent. David Pugmire, for example, suggests the following: if one fails to resist a usurping desire, when "one has made one's best effort to resist," then one is unable to resist the desire (Pugmire, 1982). Pugmire argues that if an agent forms a best judgment not to do A and still acts against his best judgment, then there is nothing more the agent could have done to resist the desire to do A, since, for Pugmire, forming a best judgment is making the best effort to resist the desire. (b) Gary Watson defends Plato's theory by arguing that the unexercised ability hypothesis should be rejected. He claims that no plausible explanation for a weak-willed agent's failing to exercise his resisting ability can be "consistent" with the presupposition that the agent performs a weak-willed action (Watson, 1977). Pugmire and Watson undertake to develop arguments in line with the platonic view that akratic actions are unfree. I will argue that both attempts are unsuccessful.

III. Pugmire's Skepticism

Pugmire begins by pointing out an important feature that akratic and compulsive actions share. He claims that, when an agent performs an akratic action (Pugmire, 1982: 188),

What the agent did just because he wanted ('for no good reason') is also something he then deemed worse and not to be done. This, however, is uncomfortably close to what specifies compulsive action, which are unfree.

Even though this similarity may be granted, there is an apparent distinction between akratic and compulsive actions: the two cases involve different ability conditions of the agent. When an agent acts compulsively, he is utterly *unable* to help himself, whereas an akratic agent may have the ability to resist the temptation.

Pugmire acknowledges this commonsensical distinction, but argues that it should be rejected. On his view, when an agent acts against his best judgment, he is as unable as a compulsive agent to prevent himself from doing so—when one acts on a desire to take an akratic alternative, one acts on a desire which is irresistible. Pugmire defends this view as follows:

As everything stood the desire is as good as irresistible by him then. For [this desire] did defeat his best effort: he put himself through a deliberation that opened him, as much as anything in his power could, to what he was doing, and he reached a dissuasive all-things-considered value-judgment, resolved and set himself against what he then did anyway What more could he have done? However, if the available resources for resistance failed, it would be arbitrary to insist that the desire was resistible on the occasion and his action clearly voluntary (pp.188-189).

The main line of Pugmire's argument is that in forming a best judgment an agent has tried his best to take the preferred option; so, when the desire usurps and defeats his effort, there is nothing more the agent

could have done to resist; thus, the agent's failure to resist indicates that he does not have the ability to resist. It should be noted that Pugmire's objection does not rest on the assumption that, given that S has the best reason not to do A, if S fails to resist the desire to do A, then S does not have the ability to resist. His argument embodies a strengthened version of this assumption. What he argues seems to be this:

Pugmire's Argument

- (P1) If S forms a best judgment not to do A (on a careful deliberation) and still acts against his best judgment, then there is nothing more S could have done to resist the desire to do A.
- (P2) If there is nothing more S could have done to resist the desire to do A, then S has made his "best effort" to resist the desire to do A.
- (P3) If S forms the best judgment not to do A, and S has made his "best effort" to resist the desire to do A, but still acts against his best judgment, then S does not have the ability to resist the desire to do A.
- (C) Therefore, if S forms a best judgment not to do A (on a careful deliberation) and still acts against his best judgment, then S does not have the ability to resist the desire to do A.

This argument is immune to the objection raised against the principle that one's failing to do a certain thing entails that one does not have the ability to do it. (P3) seems to be a modified, correct version of this problematic principle. However, the problem with Pugmire's argument is that (P1) is false. It is generally not the case that, in forming a best judgment, one has tried all he could have done to resist the temptation. In other words, one's forming a dissuasive best judgment on the basis of his careful deliberation does not constitute his "best effort". Let us consider an example to illustrate these points.

Mary loves chocolate chip cookies, but she is currently on a diet because she has some medical problems and her doctor has asked her

to lose weight. Now, after dinner, there are cookies on the table, and Mary desires to have some of them, but she remembers that she is on a diet. Mary thus has both the reason to eat and the reason to refrain from eating the cookies. After weighing the two options, Mary forms the best judgment that refraining from eating the cookies is best, all things considered.

Furthermore, Mary knows from past experiences that whenever she drinks a big glass of ice water her desire to eat cookies is generally weakened. On the other hand, Mary may indulge herself in focusing her attention on the pleasant result of eating the cookies, and she knows that this sort of attending usually makes the option of eating cookies more appealing. Thus, there are two things that Mary can do to increase or decrease the attraction of the akratic alternative. Mary has, let us say, two *strategies*: the strategy of drinking ice water and the strategy of focusing her attention on the pleasure of eating. The former will increase Mary's motivation to form an intention to act in accord with her best judgment, and the latter will increase Mary's motivation to form an intention to take the akratic option. Both strategies can affect the attraction of an option without altering the fixed evaluation of the option. However, as the story goes, Mary does not drink the glass of water and lets her attention dominated by the imagined pleasure of eating cookies,⁷ and, as a result, eats the cookies, knowing that she is acting against her best judgment.

In sum, the following explanation of Mary's akratic action seems adequate. Mary eats the cookies akratically because (i) Mary desires to eat the cookies, (ii) but she forms the best judgment that refraining from eating the cookies is best, all things considered, (iii) Mary can motivate herself to act on the best judgment either by drinking a glass of water, or by avoiding focusing her attention on the pleasant result of eating cookies, (iv) but she does not, and so (v) her desire for the cookies gains

the greatest motivational strength, and she eats the cookies.

Now we may examine what is wrong with Pugmire's skeptical argument. The first premise of Pugmire's argument, (P1), entails that if Mary eats the cookies akratically, there is nothing more Mary could have done to resist the temptation, in addition to her forming the best judgment on a careful deliberation. However, the fact seems to be that, as I shall argue, there is *something* Mary could have done to resist the temptation.

As described above, Mary has at least two "strategies":

Strategy A: distracting her attention from the imagined pleasure of eating the cookies,

Strategy B: drinking the glass of ice water.

There are three points worth noting. First, to exercise either of the two strategies will increase Mary's motivation to form an intention to act in accord with her best judgment (or, will decrease Mary's motivation to form an intention to take the akratic option). Second, strategies A and B are *open* to Mary.⁸ Third, Mary knows both the above two points, that is, Mary knows the functions of the strategies, and she knows that these strategies are open to her. Given all these, it seems plausible to say that these two strategies are within Mary's ability to exercise. So, since Mary does not exercise these abilities, there is something more Mary could have done to resist the temptation.

So, (P1) is false. Mary may have made her best effort in deliberation, but clearly she did not make her total best effort, which requires her exercising one of her strategies. Thus, one of the conditions involved in the antecedent of (P3) is not satisfied, and therefore the conclusion, that the agent does not have ability to resist the desire, does not follow. Pugmire's argument should be dismissed.

IV. The Unexercisability Argument

There is however a complication involved in the above objection to Pugmire's argument. In establishing that there is something Mary could have done to resist the temptation, I have asserted that strategies A and B are open to Mary. Pugmire might object to this assertion. In particular, he might claim that Mary has the strategies but she cannot initiate them. The idea, presumably, is that Mary's desire to take the cookies is greater in strength than her motivation to act on her best judgment, and must *also* be greater than her motivation to exercise the strategies; if so, given that one always acts on the option which has greater motivational strength, Mary cannot initiate her resisting strategies, and therefore the strategies are not within her ability. Accordingly, it is sensible to say that Mary has tried her best effort when she forms her best judgment, and thus Pugmire's argument revives.

The point that an akratic agent cannot initiate his resisting strategy has been implicitly made by many who are skeptical about free weak-willed actions, but no explicit argument is offered. In the following, I will try to construct an argument called the *unexercisability argument*, which I hope recapitulates the intuitions that underlie this skeptical point.

Suppose that an agent S has a desire, D, and a best judgment, B, where B and D favor incompatible actions, and S acts on D against his best judgment. S does have a resisting strategy, E; S may try to exercise E in some situation in order to resist some desire. Assume that S succumbs to the temptation because she did not exercise E. The question is, can S exercise E? The following conditional is evident:

- (IE) If S can initiate his exercising strategy E to resist D, then it is possible that S's desire to initiate E has greater motivational strength than the motivational strength of D.

For the sake of clarity, let us symbolize the consequent of (IE) as: it is possible that

$$MS(E) > MS(D).$$

Since S's motivation to exercise E is derived from S's motivation to follow B, the motivational force to exercise E is at most equal to the motivational force to follow B, that is,

$$MS(B) \geq MS(E).$$

But, weakness of will is a case in which one's desire to take the akratic option has greater strength than one's desire to act on his best judgment. So, it is obvious that

$$MS(D) > MS(B).$$

From $MS(B) \geq MS(E)$ and $MS(D) > MS(B)$, we can derive

$$MS(D) > MS(E).$$

In other words, from the two assumptions, we can infer that in all cases of weakness of will, it is impossible that S's desire to initiate E has greater motivational strength than the motivational strength of D. Given (IE), it follows that S cannot initiate his exercising strategy E to resist D.

The flaw with the unexercisability argument lies in the claim that "since S's motivation to exercise E is derived from S's motivation to follow B, the motivational force to exercise E is at most equal to the motivational force to follow B." The consideration underlying this claim is that an agent's motivation to initiate his resisting strategy is based on his best judgment, and if his best judgment itself is not strong enough to overcome the desire, neither is his desire to initiate the strategy. This principle can be more clearly formulated as below:

In a temptation, if one's motivation to act on the best judgment has less strength than one's akratic desire, then

one's motivation to initiate one's resisting strategy also has less strength than the desire.

This principle is false because it neglects the case of *strength* of will. In a case of strength of will, an agent is tempted, and his disposition to act on his best judgment is overcome by his desire, but he manages to exercise his resisting ability and eventually acts according to his best judgment. A case in point is that Ulysses chained himself in order to resist the temptation of the Sirens. Note that a case of strength of will is similar to a case of weakness of will in that in both cases the agent is tempted and his disposition to act on his best judgment is overcome by the motivational strength of his desire. The difference between the two cases is that in the former situation the agent successfully exercises his self-control, but fails to do so in the latter. So, the existence of the cases of strength of will indicates that it is possible that one's motivation to act on the best judgment has less strength than one's akratic desire, while one's motivation to initiate one's resisting strategy has greater strength than the desire. Therefore, the unexercisability argument fails.

To repeat, the defect of the unexercisability argument is that it over-emphasizes the relation between best judgment and self-control. It should be agreed that self-control results from best judgment, but the strength of the former is clearly not determined solely by the latter. Alfred Mele's view about what he calls "the paradox of self-control" is pertinent to our discussion here.⁹ His theory seems to provide a more systematic account to support the above objection to the unexercisability argument. The basic assumption of his theory is that the motivational strength of a desire is determined by its "positive motivational force" and "negative motivational force." To be more specific, the total motivation strength of a desire, MS, in my terminology, is the result of deducting its negative motivational force (NMS, henceforth) from its positive motivational force (PMS). Given this, I derive the following principle:

Of any two desires with equal PMS, the one with less NMS is stronger in MS than the other.¹⁰

Using Mele's notions, I will reconstruct the above objection to the unexercisability argument as follows. The positive motivational force of following one's best judgment is the same as the positive motivational force of exercising the resisting effort, that is, in our terminology,

$$\text{PMS(B)} = \text{PMS(E)}.$$

It is worth noting here that this claim can be understood as the view that since an agent's motivation to initiate his resisting strategy is based on his best judgment, they share the same positive motivational strength.¹¹

Furthermore, since the usurping desire, D, directly opposes the best judgment but not the resisting effort, it is obvious that

$$\text{NMS(B)} > \text{NMS(E)}.$$

This is the central thesis of Mele's theory. His view is that one is more bothered by his usurping desire when he tries to act on his best judgment, than when he is initiating a resisting strategy. This point may be illuminated by the Mary example. Mary may form two intentions: the intention not to eat the cookies and the intention to drink the glass of water before her. Mary's forming these two intentions would be disturbed by her strong desire to eat the cookies, but in the latter case the disturbance should be less: given her strong desire to eat the cookies, it seems difficult for Mary to form the intention not to eat them, while, comparatively, it should be easier for Mary to form the intention to drink the glass of water, since forming this intention is not *directly* incompatible with Mary's strong desire to eat the cookies. So explained, Mele's view seems true.

Given $\text{PMS(B)} = \text{PMS(E)}$, $\text{NMS(B)} > \text{NMS(E)}$, and the principle about the total MS, it follows that

$MS(E) > MS(B)$.

Thus, one's motivation to initiate one's resisting strategy may be greater in strength than one's motivation to follow one's best judgment. It follows that one of the premises involved in the unexercisability argument, that the strength of one's motivation to initiate one's resisting strategy is *at most equal to* the strength of one's motivation to follow one's best judgment, is false. So, the unexercisability argument is rejected. There is no good reason for claiming that a weak-willed agent cannot initiate his resisting strategy. Consequently the Mary example remains a successful objection to Pugmire's argument.

V. Watson's Skepticism (I)

Let me briefly review what I have so far discussed. The platonic argument against free weak-willed actions fails because it is based on the false principle that if someone fails to resist a temptation, then he has no ability to resist the temptation. Pugmire's argument contains a plausible version of this principle, but fails to establish that akratic agents have tried their "best efforts." The moral to be learned from the failure of the two arguments is that, as the example about Mary shows, the fact that an agent fails to resist a temptation does not exclude the possibility that the agent has, but does not exercise, the resisting ability. A more plausible form of the skeptical argument should therefore avoid resting on the false principle, and deal with the possibility of an unexercised resisting ability.

Gary Watson's argument seems to satisfy these conditions (Watson, 1977). He acknowledges that the hypothesis that an agent has an unexercised resisting ability is consistent with the fact that the agent fails to resist a temptation, but argues that this hypothesis should be rejected. He claims that there can be *no* plausible explanation for a weak-willed

agent's failing to exercise his resisting ability such that it is consistent with the presupposition that the agent performs a weak-willed action.

To illustrate Watson's contention, recall that in explaining Mary's akratic action, it is asserted that Mary can motivate herself to act on her best judgment either by drinking a glass of water or by avoiding focusing her attention on the pleasant result of eating cookies, but *she does not*. Watson's argument is directed at this sort of explanation. He asks, granting that an akratic agent has resisting strategies which, if exercised, might be sufficient to resist the desire, why doesn't the agent exercise it? Watson thinks that if we press harder on this question, we will find that what the explanation involves is incompatible with the fact that the agent performs an akratic act. To see this, Watson asks us to imagine a case of a weak drinker who has the capacity to resist drinking. Watson so remarks:

[T]he woman judges that she should resist and therefore, by her own lights, has sufficient reason for exercising her alleged capacity. What might explain her not doing so? There seems to me only two possible explanations. (1) She *chooses* not to. (2) Her *effort* to resist is culpably insufficient. Both of these explanations will be found inadequate (p. 336).

Watson's conclusion is that, "Given her strong motive for making an effort, and in the absence of a special explanation for her not making it . . . , we are entitled to conclude that the person was unable to resist" (p. 337).

Watson argues that the possibility of (1), that the akratic agent chooses not to exercise her resisting ability, should be excluded, because "the notion of choice . . . involves the notion of applying one's value to the perceived practical option" (p. 336). Watson explains as follows:

[I]t is of course generally true that one may choose not to exercise some capacity that one has. But the capacity of self-control is special in this respect. For the capacity of self-control involves the capacity to counteract and resist the strength of desire which are contrary to what one has chosen or judged best to do. The weak drinker's failure to resist her desire to drink is a failure to implement her choice not to drink. To choose not to implement this choice would be to change her original judgment, and the case would no longer be a case of failure to implement judgment (pp. 336-337).

Watson's view revealed in this quotation is not very clear. It appears to say the following:

- (i) If S judges that it is best not to do A, then S chooses not to do A.
- (ii) If S chooses not to exercise his ability to resist the desire to do A, then S chooses not to implement his choice not to do A.
- (iii) If S chooses not to implement his choice not to do A, then it is not the case that S chooses not to do A.
- (iv) Therefore, if S chooses not to exercise his ability to resist the desire to do A, then it is not the case that S judges that it is best not to do A.

It follows that when S does A, S does not act against his best judgment. Now, given that the weak drinker chooses not to exercise her ability to resist and, as a result, drinks akratically, if (iv) is true, it follows that the weak drinker does not judge that it is best not to drink, and therefore she does not act against her best judgment. In other words, if the argument is sound, the explanation that the agent choose not to exercise her resisting ability is *inconsistent* with the fact that she performs a weak-willed action. It is due to this sort of inconsistency

that Watson claims that the explanation cannot enter into an adequate explanation of weakness of will.

Some philosophers find this argument weak. According to them, Watson assumes too tight a relationship between best judgment and choice. Walker, for example, holds that, "Watson rejects the explanation that the agent chooses to perform the akratic action, but this rejection is based on a controversial claim that choice must follow better judgment" (Walker, 1989: 656). For another example, Mele claims, "The argument about the weak drinker's failure to resist her desire to drink depends upon a disputable identification of better judgment and choice" (Mele, 1987: 28).

It should be clear that the two authors criticisms are directed at premise (i) of the above argument. However, in my opinion, Watson's argument can be formulated in a way in which (i) is dispensable. Assuming S judges that to do A is better than to do B, and S has a strategy which, if exercised, can prevent him from succumbing to the temptation of doing B, we can revise the argument as below:

- (P1) If S chooses not to exercise his resisting strategy, then S evaluates that not to exercise the strategy is better than to exercise it.
- (P2) If S evaluates that not to exercise the strategy is better than to exercise it, then S evaluates that to do B is better than to do A.
- (C1) So, if S chooses not to exercise his resisting strategy, then S evaluates that to do B is better than to do A.

This form of Watson's argument focuses on the relation between choice of self-control and best judgment, rather than the relation between choice and best judgment. This modified argument has two advantages. First, it avoids the problematic premise involved in the previous version of the argument. Second, it accommodates Watson's view that the capacity for self-control is distinctive from other abilities, with respect to one's

choice.

Yet, this modified argument is objectionable as well. Both premises are questionable. To see the difficulty with (P1), consider again the example of Mary. Mary knows, from past experiences, that focusing her attention on the pleasant result of eating cookies usually makes the option of eating cookies more appealing. Mary can, of course, choose to be in a state of attending to certain things, knowing that she had better not, for there might be some undesirable consequences, but nevertheless indulges herself. This case of choosing to indulge oneself is certainly possible. In fact, this sort of self-indulgence seems to be involved in most (if not all) cases of weakness of will.¹² Now, if Mary chooses to indulge herself in attending to the cookies, then it is not the case that she judges that to attend to the cookies is better than not. For if it is, then Mary's focusing her attention on the cookies cannot be correctly described as a form of self-indulgence.¹³ Furthermore, Mary's strategy A to resist eating the cookies, as explained, is to distract her attention from the cookies. Thus, to say that Mary chooses to focus her attention on the cookies is to say that Mary chooses not to exercise her resisting strategy. To sum up, Mary chooses not to exercise her resisting strategy, and it is not the case that she judges that not to exercise her resisting strategy is better than to exercise it. Thus, (P1) is false.

(P2) claims that when one judges that not to exercise one's resisting strategy is better than to exercise it, one no longer maintains one's best judgment. There is a type of case in which (P2) is obviously false. Imagine that the weak drinker does not know the fact that her acting on the best judgment requires her exercising the resisting strategy. It is conceivable that she may judge that it is best not to drink, and has a strategy to resist the temptation to drink; but she finds that the temptation is so weak that she need not exercise the strategy, and so she evaluates that not to exercise the strategy is better than to exercise it. That is, based

on the consideration about the cost of exercising a resisting strategy (e.g., it will be a waste of time and energy to do so), one may rationally judge that it is better not to exercise the strategy, while maintaining one's best judgment. So, (P2) is false. Both premises are questionable, and therefore Watson's argument is unsound. It is very unlikely that when one chooses not to exercise his resisting strategy, he changes his best judgment. There seems no such tight connection between choice of self-control and best judgment.

It is worth noting that, regarding the above counter-example to (P2), since the consequence of the example is that the agent fails to resist, the weak drinker's belief about the effort required in resisting is false. The agent *misjudges* what it requires to resist the temptation, and therefore judges that not to exercise the resisting strategy is better than to exercise it, but nevertheless maintains his best judgment. In other words, (P2) is false when an agent's misjudgment is involved. I will discuss the misjudgment case in the next section because Watson contends, in his second argument, that misjudgment and weakness of will are incompatible.

VI. Watson's Skepticism (II)

Watson's argument denying the possibility that an akratic agent chooses not to exercise her resisting ability has been refuted. In his second argument, Watson claims that the explanation offered in (2), that an akratic agent does not make a sufficient effort, is not an adequate explanation for weakness of will. Watson first points out that (2) assumes that an akratic agent is able to make the requisite effort to resist the temptation. Further, he claims the following:

If effort of a certain kind and degree is necessary to successful resistance, it will be true that the drinker is able

to resist only if she has the capacity to make an effort of that kind and degree to resist... Our focus is thus shifted to her failure of effort, and everything now turns on why she does not make it (p. 337).

Why doesn't the agent make the requisite effort? Watson raises two possible explanations and rejects both. The first possible explanation is that the weak drinker does not make the effort because she thinks that making the effort is not "worth" it.

The explanation cannot be that making the effort is not thought to be worth it ... [I]mplicit in the judgment that it is best not to drink is the judgment that it is best to resist contrary desires. If the drinker really judges that it is not worth that much effort, she either changes her mind or originally only made a conditional judgment of the form: it is best not to drink unless not doing so requires too much effort (p. 338).

Watson apparently thinks that it is inconsistent for the weak drinker to judge both that it is best not to drink, and that it costs too much to exercise the strategy to resist drinking. He concludes that when the agent *really* judges that it is not worth that much effort, the agent must have changed his best judgment. It follows that, when the agent drinks, he does not act against his best judgment, which is contrary to the presupposition that the agent performs a weak-willed action.

The second possible explanation that Watson intends to rule out is that the weak drinker "misjudged the amount of effort required" (p. 338). The drinker might have underestimated how much effort is required for successful resistance, and for that reason she did not make a sufficient effort. Watson contends that it is unclear what kind of misjudgment is involved here, and "even if misjudgment were involved, that would be a

different fault from weakness of will" (p. 338).

The misjudgment case mentioned earlier, as we shall see, plays an important role in our present discussion of Watson's two objections. The first objection is based on the view that there is an inconsistency involved in the weak drinker's judging both that it is best not to drink, and that it costs too much to exercise the strategy. At the first sight, this claim is obviously false, for it is clear that an inconsistency is involved only when a belief condition is added, namely,

that the agent believes that resisting the desire requires his exercising the strategy.

For, if the agent believes that he can act on his best judgment without exercising his resisting strategy, he would of course think it is not worth the effort, while consistently holding his best judgment.

This is a clear defect of Watson's first objection, but to refute the objection in this way sets the stage for Watson's second objection. For if the agent does not make the requisite effort because he falsely believes that resisting the desire does not require his exercising the strategy, then his failure to resist is due to his misjudgment about the amount of effort required, which according to Watson is "a different fault from weakness of will." By this he means that if an agent's misjudgment is involved it is not a case of weakness of will.

If this is correct, then we should assume that the above belief condition holds, so far as weakness of will is concerned. That is, a weak-willed agent must realize the requiring relation between his acting on his best judgment and his exercising his strategy. But, then, why doesn't the agent exercise his strategy? Watson implies that the only plausible explanation is that the agent judges that making the effort is not worth it. Then, we fall back to Watson's first objection.¹⁴

To examine Watson's view, let us modify the example of Mary.

Suppose that Mary will be sure to act on her best judgment if, and only if, Mary exercises both strategy A and B to the fullest extent. Moreover, Mary fails to resist the temptation, because she does not make the requisite effort in exercising her resisting strategies, that is, either she does not exercise both strategies, or she exercises them but not to the extent that is sufficient to resist the temptation. Now, the question concerning us is, can it be the case that Mary fails to make the requisite effort because she judges that it is not worthwhile? The answer should be yes. As I have argued in the previous section, an agent thinks it is worthwhile to make the effort only if the agent believes that her acting on her best judgment requires that she make the effort. Her holding the best judgment alone does not entail that she thinks it is worthwhile to make the effort. Accordingly, if Mary does not know that to resist her desire to eat the cookies she has to make the amount of effort required, it is possible that she does not think it is worthwhile to make the effort. The first objection is rejected, but the second objection surfaces.

To say that Mary does not know she should exercise the strategy to the extent required to resist the temptation is to say that Mary misjudges the effort she should make in order to act on what she thinks is best. If so, Mary's action is based on a misjudgment. However, it appears that, when a person displays weakness of will, his problem is not that of judgment but rather of his acting: he does not act in an appropriate way. As Watson claims, misjudgment is a different fault from weakness of will. It follows from this that, if the explanation of why Mary fails to make the effort to exercise her resisting strategies necessarily invokes her misjudgment, it is not an adequate explanation for weakness of will.

This is the main structure of Watson's reasoning. Now, we can see that Watson's skeptical argument against free weak-willed actions pivots on the following claim: When an agent's misjudgment is involved it is a different fault from weakness of will, and thus no adequate explanation

of weak-willed actions can include an agent's misjudgment as a factor. However, Watson does not try to vindicate this claim. He apparently regards this claim intuitive. Many philosophers who adopt Watson's skeptical position seem to share the same intuition.¹⁵ However, I do not find Watson's claim intuitive. In the following I will examine some possible grounds on which this claim can be justified. I think there are two possible ways to justify Watson's contentions: *the incompatibility thesis* and *the misjudgment argument*. But both, as I shall explain, should be rejected.

(a) The incompatibility thesis. It should be clear that Watson's claim that misjudgment is a different fault from weakness of will, by itself, does not constitute an objection to the view that misjudgment may play a role in explaining weakness of will. Objecting to this view requires a stronger claim. It requires a claim advancing from the difference between misjudgment and weakness of will to the incompatibility between them. At the least, the following principle must be included.

(Incompatibility thesis) The explanation that an agent does something due to his misjudgment is incompatible with the explanation that he displays weakness of will.

What would be the grounds for accepting this thesis? Intuitively speaking, the incompatibility thesis may be based on the distinction between *judgmental* and non-judgmental issues. The idea is that misjudgment is a fault about one's judgment, but weakness of will is not a judgmental fault, and therefore, if misjudgment is involved, it is not a case of weakness of will.

The incompatibility thesis is obviously too strong. The truth seems to be that misjudgment can be a factor in explaining weakness of will, when it is related to, based on, or can be explained by one's weakness. To see this, consider the Mary example. Suppose that in order not to

eat the cookies, Mary must fully exercise her resisting ability. Mary has overwhelmingly good evidence for believing that fact. However, Mary is tempted by her desire to eat the cookies so that she comes to believe, contrary to the evidence she has, that she can resist the desire to eat the cookies without fully exercising her resisting strategies. Due to this misjudgment, she does not fully exercise her resisting ability, and consequently eats the cookies. In this case, Mary displays weakness of will in succumbing to the temptation and acting against her best judgment, but her misjudgment clearly plays an important role. There seems no incompatibility between misjudgment and weakness of will. The thesis that misjudgment excludes weakness of will should therefore be rejected.

(b) The misjudgment argument. Some might want to defend the view that misjudgment excludes weakness of will by arguing that the very definition of weakness of will excludes the possibility of misjudgment. Their view can be represented by the following argument.

The Misjudgment Argument

- (P1) If an agent performs a weak-willed action, then, by definition, he knowingly acts against his best judgment.
- (P2) If an agent's misjudgment is involved, then the agent does not know well enough what he is doing when he acts.
- (P3) If the agent does not know well enough what he is doing when he acts, then he does not knowingly act against his best judgment.
- (C) Therefore, if an agent's misjudgment is involved, then it is not the case that he performs a weak-willed action.

This argument is valid, but it embodies an ambiguity. (P3) is true only if what the agent fails to know *includes* the fact that he acts against his best judgment. For it is obviously possible that an agent does not know well enough what he is doing because, for example, he misjudges the amount of effort required to resist the temptation, but when he acts, he does know that he is acting against his best judgment.¹⁶ So,

if (P3) is true, that the agent does not know well enough what he is doing entails that the agent does not know the fact that he acts against his best judgment. But then, (P2) is false, because the fact that the agent misjudges the amount of effort required to resist the temptation, of course, does not entail that the agent does not know that he acts against his best judgment. The latter fact is clearly independent from the former. In other words, the misjudgment argument suffers from the fallacy of equivocation. The range of what “the agent does not know well enough what he is doing” is ambiguous. If it includes the fact that the agent acts against his best judgment, then (P3) is true but not (P2). On the other hand, if it doesn't, then (P2) is true but not (P3). So, the misjudgment argument is rejected.

The incompatibility thesis and the misjudgment argument are both refuted. Thus, there seems no good reason to support Watson's claim that misjudgment cannot be involved in any explanation of weak-willed actions. Therefore, the hypothesis that an akratic agent has some unexercised ability to resist the temptation remains a plausible explanation of weakness of will.

VII. Concluding Remarks

The purpose of this paper, as stated in the introduction, is to defend the possibility of free weak-willed actions. This purpose has been served in the following way. The No Ability Argument, the prime argument for the skeptical view that one acts against one's best judgment only if one does so unfreely, is based on a principle: “if an agent succumbs to temptation and acts against his best judgment, then he acts on a desire which he has no ability to resist.” Three main arguments for this principle, proposed respectively by Plato, Pugmire, and Watson, have been examined. To refute these arguments, I have described the example

of Mary. This example presents a plausible case in which the agent, Mary, performs a weak-willed action, but (i) she has the ability to avoid doing it, and either (ii) she chooses not to exercise her resisting ability, or (iii) she does not make a sufficient effort to exercise her resisting ability. This explanation of Mary's action indicates that the fact that an agent succumbs to temptation and acts against her best judgment does not exclude the possibility that the agent has the ability to resist the temptation. Thus, the central principle for the No Ability Argument is false. Consequently, the skeptical position that no weak-willed action is free can be dismissed.¹⁷

Notes

- 1 This approach, according to which weakness of will is best characterized as resulting from a split between one's evaluation and motivation, is not completely novel. Indeed, some main theses of this approach are assumed, at least implicitly, by many philosophers who discuss weakness of will. But the approach as a whole has never been clearly established. This term "the motivational strength approach" is derived from Thalberg. He launches an attack on the adequacy of this approach in his paper, "Questions about Motivational Strength" (Thalberg, 1985). Thalberg however does not spell out the motivational strength approach as I do. He is concerned only with the question of how strength of motivation is fathomed, and his main target is Davidson's partition theory of mind, not the motivational strength approach in general.
- 2 (P4) is, of course, questionable. However, in this paper, I will not challenge this principle. I take the problem of (P4) to be an issue more related to freedom of will alone, than to the relation between freedom and akrasia.

- 3 The term “usurpation” is borrowed from David Pears (1982: 167). Pears however does not make reference to Plato’s view when he develops his usurpation theory of weakness of will, but that their theories share similar features seems obvious.
- 4 This platonic account can also be understood as follows. It is assumed that one has control over one’s action only if one’s will, the faculty which governs action, is guided by one’s Intellect. However, weakness of will occurs if and only if an usurpation of desire is involved: when an agent performs a weak-willed action, he is not fully integrated so that an unruly desire takes over control of intentional action and constitutes his will. To speak figuratively, the underlying phenomenon for weakness of will is that the desire defeated in deliberation usurps the throne of the Intellect and dominates the will. So, when an agent acts against his best judgment, his will is not guided by his Intellect. Given the platonic assumption about intentional action, it follows that the agent has no control over his action.
- 5 I benefit from Santas’s excellent exposition of Plato’s theory of weakness of will, Santas (1966).
- 6 So formulated, Plato’s principle does not have the initial plausibility, but, as we shall see, this principle, so to speak, points out a direction along which many important contentions are generated, and the main purpose of this paper is to examine these contentions.
- 7 I have further discussions about why Mary, given her best judgment, does not avail herself of a strategy, in sections V and VI, where I consider Watson’s argument.
- 8 This point will be further discussed and defended in the following section.
- 9 Mele, 1987: Chapter 5 and 6, especially p. 63.
- 10 It is equally derivable that of any two desires with equal NMS, the one with more PMS is stronger in MS than the other.
- 11 There could be, of course, other positive factors affecting E, that is, there may be independent reasons to do E. For example, as Richard Feldman

points out to me, it is possible that S's character is such that he values highly his displaying the strength of will. If so, S has an extra reason to do E, in addition to his considerations that support his best judgment. In other words, it is possible that the positive motivational strength of S's doing E is greater than S's motivation to follow his best judgment. So, the principle should be modified as follows: The positive motivational force of following one's best judgment is at most equal to the positive motivational force of exercising the resisting effort. That is, $PMS(B) \leq PMS(E)$. This may be granted. But it should be clear that this modification does not weaken the argument, and in fact strengthens it. The idea is that given $PMS(B) \leq PMS(E)$, even if B and E share the same negative motivational strength, it follows that the total motivational strength of E is at least equal to that of B. This, in my opinion, may be an even *better* way to argue against the unexercisability argument.

- 12 It should be noted that Watson admits the possibility of weakness of will, that is, he admits that it is possible that one acts against one's best judgment. What he argues against is the view that it is possible that one acts *freely* against one's best judgment.
- 13 An anonymous referee points out that I should explain how this "indulgent choice" is possible, because he/she thinks that Watson might want to push the following question: "Why does the weak, given his clear awareness of what is rational for him to do, choose to ignore the required resisting strategy?" In my view, the main problem with Watson's argument lies exactly in the claim that indulgent choice is *impossible*. To begin with, the concept of choice simply does not imply that one always chooses what is the best. Moreover, the example of Mary is plausible enough—there is nothing outrageous for Mary to choose to focus her attention in a certain way. Regarding the question why Mary chooses to attend to the cookies and thus fails to exercise her resisting strategy, I think the proper answer is that because she *likes* to attend to the cookies.

- 14 In light of this account of the supplementary relation between the two objections, we may understand why Watson claims that there are only *two* possible explanations for a weak-willed agent's failing to exercise his resisting ability.
- 15 I think that one reason why Watson's paper wins wide recognition is that he proposes arguments against free weak-willed action along this line. The other reason may be that in the paper he successfully refutes a quite commonly accepted view, namely, that since we normally hold a weak-willed agent responsible for his action, a weak-willed action must be done freely.
- 16 It should be obvious that the knowledge condition involved in saying that one knowingly acts against his best judgment is merely that one knows his best judgment about his options, and knows that what he is doing is incompatible with his following the best judgment; to be sure, it does not require that one has perfect knowledge about everything involved.
- 17 I am grateful for comments on previous drafts of this paper from Earl Conee and Richard Feldman. I also wish to thank two anonymous referees for their helpful suggestions.

References

Mele, A.

1987 *Irrationality: An Essay on Akrasia, Self-deception, and Self-control*. New York: Oxford University Press.

Pears, D.

1982 *Motivated Irrationality*. Oxford: Oxford University Press.

Plato

1985 *The Collected Dialogues of Plato*. E. Hamilton and H. Carins (eds.), New Jersey: Princeton University Press.

Pugmire, D.

1982 "Motivated Irrationality," *Proceedings of the Aristotelian Society*
56: 179-196.

Santas, G.

1966 "Plato's *Protagoras* and Explanation of Weakness," *Philosophical Review* 75: 3-33.

Thalberg, I.

1985 "Questions about Motivational Strength," in E. LePore and B. McLaughlin (eds.), *Actions and Events*, pp. 88-104.

Walker, A.

1989 "Weakness of Will," *Nous* 23: 653-676.

Watson, G.

1977 "Skepticism about Weakness of Will," *Philosophical Review* 76:
316-339.

Williams, B.

1979 "Internal and External Reasons," in R. Harrison (ed.), *Rational Action: Studies in Philosophy and Social Science*. Cambridge University Press.

自由與意志弱性

何志青

摘要

本文討論自由的意志弱性行爲。所謂意志弱性行爲是指行爲者的行動違反了行爲者本身的價值(最佳)判斷。我認爲很多的意志弱性行爲同時也是自由的行爲。很明顯地我們有時會自由地選擇違反我們的最佳判斷而行——雖然判斷某一行爲選項就整體而言是較好的，我們仍依自己的自由意志採取另一行爲。然而，有一廣泛流傳的懷疑論卻認爲，人只有在非自由的情況下才會作出違反最佳判斷的行爲；也就是說，根本不可能有所謂的自由的意志弱性行爲。此懷疑論的主要代表人物是柏拉圖及當代的波格麥爾和華森等人。在這篇論文中，我將檢視有關此懷疑論的三個最重要論證。爲了辯護自由的意志弱性行爲之可能性，我將證明此三論證均爲無效。