



THE CLASSICAL LIBERAL INSTITUTE  
NEW YORK UNIVERSITY SCHOOL OF LAW

**WORKING PAPERS SERIES**

BEHAVIORAL ECONOMICS AND DEFICIENT WILLPOWER

SEARCHING FOR *AKRASIA*

MARIO RIZZO

DEPARTMENT OF ECONOMICS

NEW YORK UNIVERSITY

Working Paper No. 2

December 15, 2016

Do not quote or cite without author's permission.

An index to the working papers in the Classical Liberal Institute Working Paper Series is located at <http://www.classicalliberalinstitute.org/ideas-opinions/working-paper-series/>

REVISED Feb 12, 2016

Forthcoming in *Georgetown Journal of Law and Public Policy*, vol. 14.

## Behavioral Economics and Deficient Willpower

### Searching for *Akrasia*

by

Mario J. Rizzo  
Department of Economics  
New York University

#### ABSTRACT

Behavioral economists point to deficient willpower as one of the causes of breakdowns in decisionmaking processes. Individuals, it is claimed, know their optimal course of action but sometimes are unable to implement it because they lack sufficient motivation. This paper demonstrates that the presence of deficient willpower is very difficult to ascertain empirically. There are many phenomena that, consistent with a broad notion of rationality, must be isolated before deficient willpower can be identified. Philosophers have conceptually identified many of them. Behavioral economists, on the other hand, have not been rigorous in their conceptualization of willpower and therefore are unable to identify clear cases of deficiency. Furthermore, the empirical knowledge required of the behavioral economist to identify willpower deficiency are greater than can be satisfied in a public policy context. Therefore, deficient willpower is a poor explanation of behavior and attempts to ameliorate it are inappropriate for public policy.

Philosophers and behavioral economists generally treat weakness of will (*akrasia*) or deficient willpower as a failure of full rationality. If a rational individual judges that, all things considered, the advantages of an option *a* exceed those of *b*, that individual will choose *a* over *b*. The agent's evaluative judgment and his motivation are perfectly aligned. For the akratic individual the situation is otherwise. He *knows* the better option but *chooses* the worse.

Deficient willpower is real phenomenon or, at least, ordinary people perceive it to be. Nevertheless, it is not easily distinguished empirically from other kinds of behavior. The

philosophical literature appeals to hypothetical stories in which all of the relevant information is assumed. Then it goes on to distinguish various cases of behavior that are or are not instances of weakness of will. In the course of these discussions conceptual requirements for weakness of will have been formulated and many useful distinctions have been made. On the other hand, behavioral psychologists and economists, anxious to get on with the job of testing hypotheses about willpower, have not done the necessary work to examine the phenomenon precisely. For example, how do we empirically distinguish cases of deficient willpower from those in which the individual believes that the balance of advantages between two courses of action is ambiguous? And even where the individual has made an all-things-considered judgment that  $a$  is better than  $b$  what exactly is that extra something which is necessary for the individual to choose  $a$  over  $b$ ? Why does an individual need motivation beyond the net benefits of the superior option? These questions are hard to answer without greater precision (and maybe even with).

#### The Role of Deficient Willpower in Paternalism

Before proceeding with a substantive analysis of willpower, it is useful to state why this issue is important for behavioral economics. One of the unique contributions of behaviorally-based new paternalism is the conceit that the paternalist is not forcing, incentivizing or nudging the individual to do anything he does not really want to do. Instead the individual is being moved in the direction of his true or underlying preferences. Of the various factors that could make it difficult for an individual to satisfy those preferences weakness of will is supposed to be one of the most important and widespread. The prima facie evidence for this is strikingly superficial. It consists of the observation that we all have many long-term goals that we would like to satisfy but are often tempted by the pleasures of the here-and-now. The dieter wants health but is tempted by cake. The athlete wants endurance but is tempted by relaxation. As stated, these could simply be statements reflecting the scarcity of resources and the necessity for intertemporal tradeoffs. But let this ambiguity pass for the moment.

The new paternalism is rooted in the idea of fundamental decisionmaking failure. This failure may be caused by anyone of several factors, most notably cognitive and preference biases and

lack of willpower. Due to these the agent is (sometimes or often or usually) unable to manifest his true underlying preferences in action. This is the critical claim. Without it, the entire rationale for behavioral paternalism collapses. In other work I have dealt with the biases.<sup>1</sup> Here I wish to concentrate on lack of willpower. At the outset, I want to be clear that my argument is not that decisionmaking failures due to insufficient willpower never occur. I am not advancing the argument that weakness of will is impossible. However, I am saying that behavioral economists greatly underestimate the difficulties in identifying this behavior empirically. Therefore, the real-world incidence of weakness of will is unknown.

The reason that deficient willpower has all too often become the go-to explanation for a certain class of behavior lies partially in the character of economic method. Most behavioral economists accept, as we shall see, methodological strictures that rule out the possibility of alternative explanations of the observable behavior. Furthermore, when we understand just what the concept of weakness of will entails it will become clear that it is not so straightforward and is, in fact, a complex phenomenon whose necessary conditions are not easily observed.

## The Plan

In Part One and Part Two I set up the analysis by investigating the constitutive requirements of weak-willed behavior and the role of willpower as a motivational force. In Part Three I discuss true preferences or the rational revision of plans as alternative standards against which we may be able to discern deficient willpower. In Part Four I show how the analysts' own value judgments may contaminate the assessment of a less-than-rational deviation from previous the agent's previous intentions. This is contrary to a basic principle of new paternalism that the agent's underlying valuations should rule. Part Five reformulates the issue of deficient willpower as an allocative choice on a par with the bias-free judgments people can be hypothesized to make when they are purely akratic. Part Six is a critique of both standard and

---

<sup>1</sup> See, for example, Mario J. Rizzo and Douglas Glen Whitman, The Knowledge Problem of New Paternalism, 2009 *Brigham Young University Law Review* 905 and Douglas Glen Whitman and Mario J. Rizzo, The Problematic Welfare Standards of Behavioral Paternalism, 6 *Review of Philosophy and Psychology* 409 (2015).

behavioral economists insofar as they each rule out on methodological grounds the absence of decisive or final judgments. I emphasize the idea of a process of tentative decisionmaking. In Part Seven, I show that neither regret nor self-constraining behavior offer much hope for the paternalist agenda. The final part offers a summary and conclusions.

## I. What is Weakness of Will?

Alfred Mele draws our attention to three constitutive requirements of strict or pure akratic action:<sup>2</sup>

1. It presupposes that the agent acts against his *decisive* judgement that, say, *a* is better than *b*. An agent who is uncertain, confused or finds that the relative benefits of two objects of choice are ambiguous is not suffering from deficient willpower – he is not an akratic agent. He is simply undecided.
2. The irrationality<sup>3</sup> of deficient willpower presupposes that the decisive evaluation is itself rational. In other words, it requires that this judgment is not significantly contaminated by the agent’s cognitive or preference biases. If it were, then acting against this judgment would not be objectively irrational. From the paternalist welfare perspective it would not run contrary to the agent’s “true preferences” and wellbeing.<sup>4</sup>
3. The evaluation must be all-things-considered; it must take all of the relevant factors into account and not simply be a partial evaluation across a subset of relevant dimensions. It is perfectly rational for an agent to say *a* will make him healthier than *b* and yet because of other evaluational criteria he chooses *b* rather than *a*.

The purely akratic agent embodies a curious combination of attributes. He is able to trade off many characteristics of choice-objects and make an overall comparative evaluation. He is completely decisive in that judgment; there are no ambiguities. And what is more: He is not plagued by biases or irrationalities, *except* for what is arguably the most important limitation of

---

<sup>2</sup> Alfred Mele, *Motivated Irrationality*, *The Oxford Handbook of Rationality* 240, 241(OUP, 2004).

<sup>3</sup> Increasingly behavioral economists are using the term “boundedly rational” instead of “irrational.”

<sup>4</sup> See Richard H. Thaler and Cass R. Sunstein, *Libertarian Paternalism*, *93 AEA Papers and Proceedings* 175 (2003). (“We intend ‘better off’ to be measured as objectively as possible.” at 175).

practical rationality. This is the gap between evaluative judgments and motivation to act. He knows the better and yet he does the worse. However, the weak-willed agent is still purposeful. He has not become a machine that simply responds automatically to stimuli. From a motivational perspective, he is responding to the continuing attraction of the lesser valued option. Therefore, the akratic action is still responsive to an evaluative judgment, just not the most highly ranked value.

The behavioral economics literature departs somewhat from the pure case of weak-willed action. The pure case is a synchronic, single period case. The agent has an all-things-considered definitive preference at  $t_0$  for an action which he fails to do at  $t_0$ . Behavioral economists usually analyze the multi-period case of adhering to a plan of action that is supposed to be implemented over time.<sup>5</sup> The agent plans at  $t_0$  eat fewer calories over future periods of time ( $t_1$  to  $t_n$ ). However, tempted by cake at one or more of the subsequent time periods, he fails to adhere to his plan. This case raises a very important question at the outset. Is the plan still viewed as the best course of action even as the agent departs from it? Unless that is so, the agent is not weak-willed; he has simply changed his mind. Weakness of will is a special kind of departure from the initial plan – one that does *not* involve a change in the underlying evaluative judgment.

The observer is faced with the task of differentiating between a change in evaluation at the level of the plan and an akratic abandonment of the plan in its implementation. There are two main clues to what is going on. First, if the individual continually “announces” a plan and then abandons it, we may want to conclude that he has a weak will. We might also conclude that he is not certain the plan is best. Second, if the individual feels true regret at not implementing the plan, he may be weak-willed. Of the course, the issue will be whether we are seeing actual regret or simply a disappointment that the putatively most preferred action is so costly. We shall discuss these possible confounds later in more detail.

## II. What is Willpower?

---

<sup>5</sup> One of many examples is Richard H. Thaler, *Misbehaving: The Making of Behavioral Economics* at 102-111 (2015).

Many influential psychological theorists think of willpower as a limited resource that enables agents to engage in self-control or, more broadly, self-regulation. Willpower is akin to energy which can be depleted at least in the short run by use. Previous use of this resource makes less available for subsequent use. Thus individuals can be overwhelmed by previous or concurrent other uses of this energy resource so as to make self-regulation difficult. There are both cognitive and behavioral components to self-regulation.<sup>6</sup> Forgas, Baumeister and Tice characterize self-regulation in this way:

It refers to a specific kind of change: change to bring thinking and behavior into accord with some often consciously desired rule, norm, goal, ideal, or other standard... Self-regulation is regulation of the self by the self. Thus, it means changing oneself, or some aspect of oneself, so as to conform to some idea or concept (standard). The standard may have been chosen by the self or prescribed by the social environment.<sup>7</sup>

If we link this perspective with the philosophical analysis of weakness of will it becomes clear that purposeful behavior requires both evaluative judgment *and* motivation. From the perspective of standard economics this is odd. Most economic analysis proceeds as if only the judgment component is necessary. If an individual judges that *a* is better than *b* and both are available, but only one can be chosen, she will no doubt choose *a*. *Nothing more is needed*. This type of analysis seems to work when we are dealing with most decisions people make or are hypothesized to make.<sup>8</sup> But there seem to be certain decisions that require some considerable additional motivational energy that goes beyond the motivational potential of the all-things-considered evaluation. Call that additional energy “willpower.”

---

<sup>6</sup> There is also an affective component. People may attempt to control their emotions or moods.

<sup>7</sup> Joseph P. Forgas, Roy F. Baumeister and Dianne M. Tice, “The Psychology of Self-Regulation: An Introductory Review,” in Forgas, Baumeister and Tice (ed.), *Psychology of Self-Regulation: Cognitive, Affective and Motivational Processes* 1, 4-5 (2009).

<sup>8</sup> It is quite possible that all decisions require willpower but in many cases it is so little that this additional component may be ignored.

Let us think about these cases more deeply. Before an agent can make a “rational,” all-things-considered and decisive judgment about a course of action he must be exposed to all available information. In the idealized case this means:

...relevant beliefs which are part of the “scientific knowledge’ of the day, or which are justified on the basis of publicly available evidence in accordance with the canons of inductive or deductive logic, or justified on the basis of evidence which now could be obtained by procedures known to science.<sup>9</sup>

From a behavioral perspective, simple exposure to this is not enough. The information, particularly that relevant to costs and benefits must be reflected in an adequate mental representation. The picture before him must be “ideally vivid.” Richard Brandt clarifies what an ideally vivid way of representing a desire, aversion or pleasure is:

I mean that the person gets the information at the focus of attention, with maximal vividness and detail, and with no hesitation or doubt about its truth.<sup>10</sup>

These criteria are a mix of objective and subjective considerations. Unless the agent’s plan is based on the best available knowledge departure from it is not necessarily welfare-reducing. From the perspective of behavioral paternalism weakness of will must be more than a purely formal defect in the decisionmaking process. The subjective considerations are important because to truly know the better, again from a behavioral perspective, requires more than purely intellectual comprehension and assent. Even after an individual makes a decisive overall judgment that *a* is better than *b*, *b* apparently still exercises a pull on the individual – to such a great extent that it may override the pull of the preferred option. Why might this be the case?

One possibility is that the less preferred option (according to the agent’s long-run plan) has an immediate visceral appeal.<sup>11</sup> The delicious cake right in front of the agent and available now seems better than the harder-to-envision slimmer physique in the future. One difficulty with

---

<sup>9</sup> Richard B. Brandt, *A Theory of the Good and the Right* 112 (1979).

<sup>10</sup> *Id.* at 111.

<sup>11</sup> George Loewenstein, Out of Control: Visceral Influences on Behavior 65 *Organizational Behavior and Human Decision Processes* 272 (1996).

this explanation is that is that the visceral appeal cannot be so strong as to turn the agent into a helpless play-thing of physical forces. The weak-willed individual is still an agent, that is, he still engages in voluntary purposeful behavior. Another difficulty is that behavioralists seem to be in the position of arguing that the inferior option is *too* vivid for a balanced judgment. If something can be too vivid so that it overwhelms the appeal of the other options, then the standard of “maximal vividness” becomes quite confused.

Emphasis on the vividness of the relevant facts pulls us in two opposing directions. Vividness of the facts and the available options is a requirement for the decisive judgment that weakness of will presupposes. The individual must *really know* the better before his doing the worse can be considered a lapse in willpower. On the other hand, in trying to explain what happens when the will is weak, behavioral economists frequently resort to arguments that are the equivalent of saying that the worse option appears more vividly in perception and thus is motivationally more powerful. *The knowledge requirement of weakness of will is at war with the explanation of how weakness of will is possible.* This intellectual balancing act seems to require a normatively correct degree of vividness to anchor the arguments. Where do we find that standard?<sup>12</sup>

This conundrum to one side, there is yet another fundamental difficulty. While Brandt and others emphasize not representing the long-run consequences adequately, the inadequacy of representation can also operate on the so-called short-run consequences. Consider the following case. An individual plans to run every morning before going to work. To do this she must rise at 6am, rain or shine, cold or warm, and run a mile or two. There are obvious health benefits in the long run, no doubt. But she makes this plan in conjunction, say, with her doctor when she is comfortable, the health benefits are vivid and she is long-run oriented. However, does she adequately – vividly – see the daily costs involved? She must get up, lose sleep and undergo a physically trying process. Why is the *ex ante* lack of vividness of this cost not to be lamented? <sup>13</sup>At the moment of decision each morning she perceives what her planning self only

---

<sup>12</sup> Can such a standard be ascertained empirically independently of what the analyst believes the agent’s decision should be? See note 14, *infra*.

<sup>13</sup> Most stories about these kinds of problems pass over this.

dimly saw. So when she departs from her plan she may be doing so because now she really knows the cost of these health benefits and they are not worth it. Or so it can be argued using the framework adduced to explain how weak-willed action is possible. The argument is perhaps more flexible than its proponents intend. This is due to the absence of an independent normative standard of vividness. Without one, if an agent does not adhere to her initial plan then it can be claimed that the long-term benefits were not vivid *enough* to have adhered. Equally, at the planning stage it could be claimed that the day-by-day costs of the plan were not vivid *enough* to have rejected the plan at the outset. This road is getting us nowhere.<sup>14</sup>

### III. True Preferences or Revising Plans?

1. In order to clarify further some of the complex issues involved in identifying true preferences, I will adapt an example discussed by Richard Holton<sup>15</sup>, but originally by the economist Thomas Shelling:

As a boy I saw a movie about Admiral Byrd's Antarctic expedition and was impressed that as a boy he had gone outdoors in shirtsleeves to toughen himself up against the cold. I resolved to go to bed at night with one blanket too few. That decision to go to bed minus one blanket was made by a warm boy. Another boy awoke cold in the night, too cold to retrieve the blanket and resolving to restore it tomorrow. But by the next bedtime it was the warm boy again, dreaming of Antarctica, who got to make the decision. And he always did it again.<sup>16</sup>

We can analyze this in two different ways. First, we can attempt, as behavioral economists do, to ascertain the boy Shelling's "true preferences." If we consider the conditions under which Schelling made his original plan to go to bed each night with one blanket too few, we see a boy under the influence of a movie (although presumably this wanes over time) and a nice warm,

---

<sup>14</sup> If the analyst is convinced on other grounds (e.g., his own value judgments) that one of the options is the best he may be able to give the illusion of an explanation that the agent did not have adequate willpower to take that option.

<sup>15</sup> See Richard Holton, "Intention and Weakness of Will" 96 *The Journal of Philosophy* 241, 252, 259-60 (1999).

<sup>16</sup> *Id.* at 252. The quotation is originally from Thomas Schelling, *Choice and Consequence: Perspectives of an Errant Economist* 59 (1984).

comfortable environment. When he is in bed, the cold which he could only imagine in the evening is now quite vivid. But yet he does not rise and get the extra blanket. Does he have strong willpower? No. The example suggests that, in fact, his desire to do what he believes to be best at the time in bed does not manifest itself in action because of what might be called a myopic bias. The immediate result of getting up and getting the extra blanket is to be *much* colder for a few minutes than he would be under the inadequate blankets even though by getting up he would then be warm for a longer period of time. The vividness of that more extreme cold prevents him from effectively abandoning the plan, at least at this time. So he exhibits, in a sense, weakness of will in not getting up. It is not weakness relative to the old plan but relative to the new deviating intention. Whether either of these weaknesses (failing to adhere to the plan or failing to violate it) is a problem from the behavioral perspective depends where his true preferences lie.

Possibly when the boy is in bed at night, experiencing the cost of training to be an explorer, he makes a contrary *plan*. The analytical virtue of this variation of the story is that we are not confronting long-term planning with momentary deviation.<sup>17</sup> We have now two opposing plans. The new plan is: Beginning tomorrow night and thereafter he will put the extra blanket on the bed and be warm at night. He will resist the call of exploration. Being an Antarctic explorer may now seem like an unrealistic fantasy. Unlike the original plan which uses his myopia as an enforcement mechanism, the new plan does not have such a mechanism of enforcement. When he is warm again, he reaffirms the original plan to train himself to endure the cold and he puts one fewer blankets on the bed. It clearly is an exaggeration to say that the boy “always did it again”; at some point any sane individual would stop and turn his attention, as Schelling actually did, to something else. Vacillation about plans has its own costs. When those costs become significant, behavior is likely to change. One or the other of the plans will be given up and the conflict resolved. But is the agent now better off? It is not obvious on either procedural grounds (the decisionmaking process) or on the substance of the preferences which plan is the better.

---

<sup>17</sup> This turn enables us to steer clear of the bias some analysts have in favor of long-term planning.

2. Holton rightly says that there is no clear solution along these lines.<sup>18</sup> He suggests a second possible way to analyze this story by abandoning the search for true preferences. This entails focusing on the reasonableness of *revising* a plan once it is adopted. Now we do not see the individual as being unable to implement his initial plan. We see him as changing his mind and revising his plan. Holton then outlines some criteria by which the reasonableness of revision can be evaluated. These are explicitly normative while the standard analysis is supposed to be a factual assessment about true preferences. The following are those criteria he proffers which are relevant for my purposes:

- “It is reasonable to have a tendency to reconsider intentions, if one believes that they will lead one to great suffering when that suffering was not envisaged at the time of forming the intention.”
- “It is reasonable to have a tendency not to reconsider intentions in circumstances that prevent clear thought, if those intentions were made in circumstances that allow clear thought.”
- “It is reasonable to have a tendency not to reconsider intentions that were expressly made in order to get over one's later reluctance to act.”<sup>19</sup>

The first case may or may not be relevant. Clearly, the boy's view of the costs involved undergoes a change at night. Did he envision that? Yes and no. He probably envisioned feeling cold but the picture in his head is not the same as the cold in his body. I would not emphasize this were it not for the behavioral economist's particular view of knowledge. To know in some analytical or abstract way is often considered “not enough” as when people do not respond to a list of illnesses that smoking may lead to: hence the “need” for alarming scenarios. On the other hand, visceral temptations are often considered prime impediments to effective willpower.<sup>20</sup> Presumably, the desire to get warm is a visceral temptation.

---

<sup>18</sup> Holton, *supra* note 15 at 252-3.

<sup>19</sup> Holton, *supra* note 15 at 249.

<sup>20</sup> See Loewenstein, *supra* note 11.

Secondly, Holton says we should not reconsider intentions made under circumstances that allow clear thought. I think I can often recognize in myself clearer and less clear thought. But lack of clarity is sometimes attributed others' thought when we disagree with the resulting action and vice versa. Suicide is a prime example. Consider, on the other hand, the reasonable warning in the New York subway system: Do not enter upon the tracks *for any reason*. And yet when someone intentionally jumps in and successfully saves another who has accidentally fallen on the tracks, we praise his behavior – decided in an instant – as a Good Samaritan act. Nevertheless, we need not go to the extreme cases here. The boy decided when warm and comfortable but under the influence of the movie. Where does the balance of clarity lie? It is unreasonable to have great hopes? Is it unreasonable to abandon these ambitions under the cold, but realistic, influence of the night?

The last criterion that the plan was made specifically with the contrary intention or temptation in mind is not dispositive. As I have argued above, clarity and vividness of contrary inclinations are at issue. From the behavioral perspective, there are degrees of “in mind.” Consider Brandt’s conception of vividness.

Thus Holton is correct to say that the solution he proposes is explicitly normative. Even at that, it does not provide us with clear guidance.<sup>21</sup> More importantly, Holton’s analytical move to changing plans rather than deviating from a plan does not produce the kind of solution that behavioral economists require. This is because there is an ambiguity in the use of the term “normative” as we shall see presently.

#### IV. Normative vs. Descriptive

Consider first that the behavioral inquiry may seem like a search for the *fact* of the matter – true preferences. However, this is not so straight-forward. True preferences are being contrasted to actual, observed or “revealed” preferences. They are more properly conceived of

---

<sup>21</sup> See criticism of Holton in Sarah Stroud, Weakness of Will, *Stanford Encyclopedia of Philosophy* 25-29. <http://plato.stanford.edu/archives/spr2009/entries/weakness-will/>

as *counter-factual* preferences. In our discussion here, these would be the observed preferences an individual would have if she had perfect willpower.<sup>22</sup> Without an independent measure of perfect willpower, I don't see that we can go very far at a purely positive or descriptive level. Behavioral economics, and the paternalistic policy prescriptions based on it, require us to look at decisions from the point of view of the agent. When people are nudged or taxed or otherwise regulated into better behavior in cases such as we are considering, it is on the supposition that they lack the willpower to implement *their optimal plan*. If their actions are not consistent with this, then Holton would shift focus instead to their actual plan and how or why it was "revised" relative to the optimal one.

If the new, preferred plan of action is determined by criteria of reasonable revision, the behavioral economist must be confident that these are reasonable in the individual's own context and not simply from the point of view of the analyst. This is the ambiguity noted above: normativity from the individual's point of view or from the analyst's?

A primary consideration for the individual is the contextual costs and benefits of changing her mind. Beginning with a bad plan it may not be optimal to turn it into a good plan. The initial plan can distort or at least affect subsequent judgment. Holton's reasonable grounds for plan revision thus seem short of what is needed.

Beyond that, changes in the framing of a decision – holding the facts constant – can reasonably produce changes of actions or plans. Actions are determined by desires, beliefs *and* the particular way the individual sees the facts. For example, identical facts about a person may lead a soldier to the view that the fellow in the distance is a fascist worthy of being killed or a fellow human being worthy of respect.<sup>23</sup> It depends on the meaning the individual gives to the facts from his perspective, which can change without new information. Thus when an individual changes her perspective, she may violate the previous plan. The problem is that changes in perspective *need not be reasonable or unreasonable*. So Holton's solution does not work. The

---

<sup>22</sup> Thaler and Sunstein, *supra* note 4 at 175.

<sup>23</sup> Frederic Schick, *Ambiguity and Logic* 1-7 (2003). Schick analyses a story George Orwell told of his involvement in the Spanish Civil War.

behavioralists' problem is quite similar. For them any change must be evaluated from the individual's "true" point of view. This putative true point of view leads us to a conception of true preferences. These must be established in order to determine whether the individual has simply changed her mind or has exhibited deficient willpower. Deficient willpower is found in the deviation of an action from true preferences. I don't think there is any alternative for the behavioral economist. But at some point in this chain, the analyst's value judgments will intervene because the agent's position is fundamentally too obscure to evaluate the adequacy of willpower as a purely factual matter.<sup>24</sup> Thus neither Holton nor behavioralism will give us the kind of solution required by the paternalist agenda. The new paternalist needs to use either the agent's criteria for reasonable plan change or the agent's true preferences as the standard for ascertaining deficient willpower. Instead, the paternalist ends up with a standard contaminated by his own evaluative judgments.

#### V. Does Willpower Respond to Incentives?

As we have said, willpower is most often viewed as a limited resource that is depleted by use. In the short-run the supply is fixed. In certain circumstances an individual may not have enough of this energy-like resource to implement a course of action or a plan that he believes is best, all things considered. In this strict depletion version, sufficient willpower is simply not there to use and hence the agent exhibits a failure of willpower. The typical experiment supporting this view is one in which some individuals are given a self-control task followed by another one. The standard result is that performance on the second task falls relative to those individuals who are given only the second task.<sup>25</sup> It seems as if the supply of willpower has been reduced by the first task thus impairing performance on the second task. But the finding, by itself, is misleading. More recent research is moving toward the view that the willpower available may be positively affected by the benefits associated with it as well as by changing ideas about its efficacy.

---

<sup>24</sup> On the general tendency for the analyst's value judgments to replace those of the agent when the latter are too difficult to ascertain see Rizzo and Whitman, *supra* note 1 at 965-968.

<sup>25</sup> Roy F. Baumeister and Jessica L. Alquist, *Self-Regulation as a Limited Resource; Strength Model of Control and Depletion*, in *supra* note 7, at 24.

Often when people do not use sufficient willpower in a given circumstances it is not because there isn't enough left. Rather, it is because they choose not to use it. For example, when individuals *know* that a subsequent self-control task will follow the initial one their performance on the initial task is lower and the subsequent task higher than when they do not know it will follow.<sup>26</sup> So willpower resources are conserved for later use. This means that at some level the use of willpower is a choice. We can extend the argument further. When people have been engaged in a supposedly depleting self-control task and are thereafter presented a task that has been incentivized by money payments, there is no reduction in performance in that second task. Furthermore, people can, to a certain extent, choose to structure tasks so as to focus on the enjoyable or rewarding aspects. If they feel such intrinsic motivation to pursue a task, the application of willpower can thereby be increased. Smokers who have been successful in resisting smoking at an initial time period exert increased effort at resisting in the next time period. One plausible explanation of this is that they now see self-control as more productive than they previously did. Similarly, when individuals are induced not to *believe* that their willpower is depleted by an initial task, their performance on the second is better. It is as if more willpower is available. In all of this, motivation is the key.<sup>27</sup>

When we think of the exercise of willpower as subject to choice rather than to fixed availability, then it is no longer appropriate to talk of individuals who do not implement the optimal plans or actions as irrational or even less than rational. It is not the case that they know what is better but are *unable* to act accordingly. There are many uses to which willpower can be put. If it is not used in a given case it is usually because there are other situations in which it is more urgently needed. Willpower, like all scarce resources, has an opportunity cost. Thus there is no lack of rationality when its exercise is affected by extrinsic incentives and intrinsic incentives including the perception of willpower efficacy. This is how we would expect rational individuals to behave.

---

<sup>26</sup> Id. at 25.

<sup>27</sup> Robert Kurzban, Angela Duckworth, Joseph W. Kable and Justin Meyers, An Opportunity Cost Model of Subjective Effort and Task Performance, 36 *Behavioral and Brain Sciences* 661, 671-73 (2013).

Behavioral paternalists would probably argue in response to the above analysis that carefully crafted policies can serve as a substitute for personal willpower or self-control. By making certain “good” options easier to choose (nudges like defaults) or certain “bad” options harder to choose (sin taxes), policies can ensure that individuals more often choose the better over the worse. However, this view does not deal with the basic issue. When people have the willpower reserve but do not use it, how do we know that they really have a *decisive* preference for a certain action or plan? I turn to this topic next.

## VI. The Abstraction of the Decisive Judgment and Preference

The preferences that are most relevant to my concerns are “intermediate preferences.” They are, by and large, the most common form of preferences. We can contrast these with hard-to-observe intrinsic preferences which are defined over objects of choice that are valued for their own sake. Intermediate preferences are for objects or courses of action that are expected to attain some goal that is not final. They are means to ends.<sup>28</sup> For example, if a person has a preference for saving a certain amount each month in a particular mutual fund, it is probably because he believes that it will be an efficacious means to achieve the end of maximizing the return on his savings. He wants this because he thinks that is the best way to have a good retirement. But having a good retirement is just a part of having a good life. So maybe he should not only make judgments about the fund and his monthly contributions but also think about spending more money on vacations while he is still young. The point is that the formation of preferences is intertwined with various factual judgments and predictions. It is rare that these judgments will be made in a decisive fashion *ab initio*.

Both behavioral and standard economists usually think of preference orderings as complete and unambiguous. An individual either definitely prefers *a* to *b* or is definitely indifferent between *a* and *b*. There is no room for undecided or ambivalent individuals. But what is the evidence for this? There is no direct evidence that people are always decisive in their judgments

---

<sup>28</sup> We might question whether there are truly any ultimate ends. The economist Frank Knight argued that what people really want are better ends. An end once attained becomes the basis for further ends. Frank H. Knight, *The Ethics of Competition*, in *Selected Essays by Frank H. Knight*, vol. 1 (“What is Truth” in Economics?), ed. Ross B. Emmett, 61-93 (1999).

and preferences. The assumption that they are is made by economists for two reasons. First, in many situations the idealization does little harm. For example, if we are trying to predict the reaction of consumers to a change in prices the definiteness with which preferences are held is not of first-order importance because we are mainly interested in the direction of change. Second, the assumption makes analysis far more tractable than if we have to deal with fuzzy preferences. Nevertheless, when we are trying to make welfare statement about people who appear to fail in their pursuit of stated or planned courses of action, it becomes incumbent on us to determine whether we have a case of deficient willpower or *lack of decisiveness in their preference judgments*.

To make a decisive judgment regarding intermediate preferences thus requires two elements of decisiveness: (1) a judgment about whether an intermediate or more nearly ultimate end – say, an incrementally more comfortable retirement is worth the cost of more (vacation) consumption today and (2) a judgment about the causal relation between the means chosen and the intermediate end – the efficiency of the means to that end. We should also note that the decisionmaker must decide facing an uncertain future. Each element of the decision will be affected by the “dark forces of time and ignorance.”<sup>29</sup>

Unfortunately, all of this is obscured by the tendency of economists to elide the difference between preferences and choices. It is unfortunate that economists sometimes use the term preference as the equivalent of choice or decision.<sup>30</sup> Only by a strict behaviorist interpretation is a mental event such as preference reducible to its outward manifestation.<sup>31</sup> In the cases we are discussing here, this means confusing the external act in its definiteness with the decisiveness of the intellectual or psychological state that produces it. Admittedly, people do

---

<sup>29</sup> John Maynard Keynes, *The General Theory of Employment, Interest and Money* 155 (1936). See also, generally, Gerald P. O’Driscoll, Jr. and Mario J. Rizzo, *Austrian Economics Re-examined: The Economics of Time and Ignorance* (2015).

<sup>30</sup> See Mario J. Rizzo, James M. Buchanan: Through an Austrian Window, 27 *The Review of Austrian Economics* 135, (2014).

<sup>31</sup> See a critique of this approach in Mario J. Rizzo, The Problem of Rationality: Austrian Economics between Classical Behaviorism and Behavioral Economics, *The Oxford Handbook of Austrian Economics* 364 (2015).

act on the basis of their preferences *at the moment of action*.<sup>32</sup> But these mental preferences need not have been decisively or definitively formed.

As I have said, the view of deficient willpower advanced by behavioral economics rests on a difference between a plan, intention or resolution to do something and a subsequent action. An individual resolves, for example, to exercise every morning but does not do so. Was her resolution decisive? Perhaps it was, but it need not have been so.

Nicholas Rescher makes a very useful distinction between two forms of reasoning: linearly inferential and dialectically cyclic.<sup>33</sup> The first is deductively valid reasoning based on premises that are certain. In this case the conclusions are certain and are consistent with each other. In the second case the premises are merely plausible or plausibly true and they may not be mutually consistent. Thus the inferred conclusions are not certain; they are simply provisional that is, subject to change as the matter is reconsidered. This is a form of dialectical reasoning:

The root idea of such reasoning is that of a multi-stage process whereby we repeatedly examine one self-same issue from different, and mutually inconsistent, points of view. It is a matter of developing a course of reasoning in several phases or “moments.” We proceed in circles or cycles, returning repeatedly to a certain issue, viewing it now in this light and now in that.<sup>34</sup>

If we think of the individual’s decisionmaking as reasoning toward one or more intermediate preferences on the basis of which he will perform actions or make plans, then it is easy to see that factual and specifically causal premises will be involved. In the case of resolutions or plans that are expected to govern behavior over an extended future, there will be considerable uncertainty about the factual premises. Even the number and scope of factors to be evaluated is not simply given. Remember that a decisive judgment is all-things-considered. Initially, the

---

<sup>32</sup> We take this as axiomatic. However, this is not a tautology. It is a commitment of “folk” (commonsense) psychology. See Rizzo, *supra* note 30 at 139-141.

<sup>33</sup> Nicholas Rescher, *Rationality: A Philosophical Inquiry into the Nature and Rationale of Reason* 83-90 (1988).

<sup>34</sup> *Id* at 83-84.

individual's judgment may not adequately encompass all of the relevant factors in his decision; these must be discovered.

Furthermore, an individual may need to imagine future states of the world and his future preferences under those states before he can settle on a plan. For example, whether a 25 year-old resolves to stop smoking may depend on what he thinks the future course of medical developments will be. He will also have to consider just how unpleasant resisting temptation might be or how much he will enjoy smoking over a period of time. A person thinking of saving more for his retirement will have to decide whether it is worth the opportunity cost of lost consumption and this evaluation might change as the individual considers different objects of consumption. It will also depend on how much he thinks he will enjoy, say, a vacation when he is less vigorous than when he is more. Or even what the likelihood is that the state will provide for him if he runs out of money in old age or how long he might live in view of family history of disease. Even if the fundamental facts before him do not change during the process of reasoning, his perspective or framing may. This is enough to cause reconsideration of a possible action or plan. A cycle of dialectical reasoning may characterize an individual's decisionmaking process *before* a decisive preference and plan is established (if it is ever established).

From a behavioral perspective, there may also be shifts in the vividness with which costs and benefits are apprehended. These shifts may also affect the decisions that are made; the individual may perceive a certain course of action as worthwhile or not as benefits are viewed more vividly or costs are viewed more vividly. However, the shifts cannot simply be psychological facts of the individual's experience. They must have a normative character if they are to do the work required by behavioral paternalists. Paternalists must say that some benefits or costs *should* be vivid and others not. This normative commitment by the analyst is crucial to the definition of true preferences.<sup>35</sup> Without it, we simply have lack of decisiveness.

---

<sup>35</sup> Whether the normative commitment of the analyst-paternalist reflects that of the agent is an open question. Consistency with the rationale of new paternalism requires that it does.

## VII. Tell-Tale Signs of Weakness of Will?

If, as I have suggested, weakness of will exists, then should there not be some unambiguous external signs? Two possibilities are regret and self-constraining behavior.

1. True regret is the thought that, given what I knew then, I should have done otherwise.

Passing over for the moment any ambiguities of what it means to “know” something, regret may be purely intellectual or it may have a hedonic or emotional component. In either form, it is a natural response to looking back at a failure of willpower. However, it is not simply a reaction to the eventuality that things turned out worse than expected. For example, I can choose to save less for my retirement expecting a higher rate of return than I got. I will then be disappointed at the outcome. I might even say “I regret not saving more.” But if I chose on what I “reasonably” believed was the best of available advice, this is not true regret but simple disappointment. There was no failure of willpower involved. I did not know the better and yet choose the worse. I suggest that people often use the word regret in an imprecise way. This is “regrettable.” But it is understandable the reasonableness of one’s beliefs, including the vividness and lack of ambiguity with which one held them, is not easily ascertained – most especially by observers or analysts, but also by the individual himself.

2. The most revealing sign that the individual expects herself to be subject to deficient willpower is the choice to constrain one’s own behavior at some future time. For example, Jane loves parties and generally has a good time but, before she is at the party, she thinks of herself as drinking too much there. So she resolves to limit her drinking at the next party. But because she believes she has deficient willpower she goes to the party with a friend who promises to stop her from drinking more than three glasses of wine. The friend acts to constrain her behavior; the friend is in effect a commitment device initiated by Jane herself. Why would Jane do such a thing? After all, if she wants to stop at three drinks, she should just do so. Jane – before the party – is not confident that she will have the strength of will to carry through on her resolution, her plan. The self-constraining behavior reveals her problem. Nevertheless, *it is simultaneously recognition that she has a problem and an attempt to solve it*. However, a paternalist might argue that if the attempt to defeat deficient willpower is unsuccessful Jane

has not been clever enough. She should have found a better way to limit her drinking. Perhaps she should not have gone to the party at all. This, of course, would have been a more costly option. After all, Jane gets enjoyment from the parties. Jane may wish to limit her drinking but not at the cost of not going to the party. The benefit of drinking less does not exceed that cost, as she perceives it before the party. Is she wrong? Does it make sense for a paternalist to try to prevent her from going or to tax her partying? The optimal self-constraining behavior for Jane in the specific circumstances in which she finds herself is far from obvious to others. Thus observing Jane's self-constraining behavior is indeed evidence that she believes that she has a self-control problem. It is also evidence that she has attempted to deal with it. If she succeeded then that is that. If she did not, it may simply be because the incurring of additional costs is not justified by the outcome. A paternalist resistant to this conclusion would then proceed to make the deficient willpower argument at yet a higher level: Jane does not have the willpower to implement effective self-constraining behavior.<sup>36</sup> And on we go.

#### VIII. Conclusions: The Confounds of Deficient Willpower

My purpose was to undermine the often-facile use of weakness of will as an explanation for decisions that seem puzzling or inconsistent to behavioral economists. A serious analysis of weakness of will recognizes the complexities and difficulties in identifying weak-willed behavior empirically. Even at first glance, the constitutive requirements are substantial. The agent must act against an all-things-considered, decisive judgment – rationally arrived at. Thus the agent must be less than fully rational insofar as he has willpower problems but perfectly rational insofar as his not-directly-observable judgments are concerned. These judgments are then treated by behavioral economists as the agent's true preferences.

The akratic agent deviates from his true preferences because he lacks sufficient motivation to take the better option. He does not lack knowledge of what the better option is. But what does

---

<sup>36</sup> Jane might not adequately constrain herself because she underestimates her willpower difficulties at the party. Presumably, she will soon find out.

it mean, from a behavioral perspective, to *know* the better? The behavioralist requires not simply knowledge in the intellectual sense but knowledge that is sufficiently vivid to motivate behavior. Without an independent measure of vividness and without a norm of ideal vividness, the whole issue of knowledge becomes quite confused. Some pieces of knowledge can be insufficiently vivid (often long-run consequences) and other pieces are too vivid (often short-run costs).

The issue of vividness runs into the concept of willpower. Too much vividness is, in effect, the explanation of deficient willpower. The inferior option is more vivid than the superior one. Thus there is insufficient motivation to take the superior. Or we can put it the other way, there is too little vividness about the superior option and the facts related to it. But vividness is crucial to the agent's *recognition* of the better option in the first place. The sharp motivation-evaluation dichotomy so important to the idea of weakness of will is consequently blurred. Knowing and being motivated are related – perhaps too much so for the intellectual comfort of behavioralists.

Standard neoclassical economists are accused by behavioral economists of the simplistic view that if an individual chooses to depart from a plan it is because she has merely changed her mind about whether the plan is the best course of action. She has decided to allocate her resources in another way according to rational principles under the constraint of scarcity. Instead, behavioral economists will sometimes argue that the real reason is that she lacks the willpower to execute her optimal plan. Psychological research, however, provides evidence that cases of deficient willpower are usually not caused by the exhaustion of a strictly limited supply, but by an allocation decision. Normally, the willpower-energy is there but the individual *chooses* to allocate elsewhere where the need is more urgent. This insight changes the nature the deficient-willpower argument for paternalism. It puts the argument, contrary to the intent of behavioral economists, on a level with the presumptively rational judgments relative to which weak will is defined. People sometimes do not choose the better option because they choose to allocate their scarce (energy) resources elsewhere. The paternalist is, in effect, saying that they ought to have allocated these resources in the case under discussion as opposed to

somewhere else. How does the paternalist know that? Does he have access to the array of self-control options before the individual?

The assumption of decisiveness in the agent's judgment – another of the constituent requirements of weakness of will – is a sometimes convenient methodological assumption but one that precludes recognizing that the agent may in the midst of a *process* of judgment-making. Cyclic dialectical reasoning can give the appearance of deviating from a settled plan where the plan is not at all settled. The tentativeness of the judgments upon which people make decisions is pushed aside in favor of the idealization of the completely made-up mind.

While regret and self-constraining behavior can be indicative of weakness of will, they are not always. Sometimes regret is not true regret but a disappointment that things did not turn out well. Self-constraining behavior does reveal that the agent thinks she has a willpower problem but it also points to a solution.

I do not deny that many of these confounds and conceptual confusions might be sorted out in the context of individual psychotherapy or analysis. This, of course, could produce many benefits for the individual. Nevertheless, *public policy* cannot be made on the basis of individual therapy. I do not see behavioral economists dealing with these issues. Until they do, we should all remain unconvinced that deficient willpower ought to be an element of public policy making.

*I am indebted to comments from participants of the symposium on the Ethics of Nudging, Georgetown University, November 20, 2015 and to Robert T. Miller (University of Iowa College of Law).*

