

# LIBERALIZING THE CONCEPTION OF RATIONAL ACTION

*Shifting the focus towards deliberation preceding an action*

**Abstract:** Rational action is commonly considered to result from a concurrent better judgment. This way we lose eye for the rational aspects of actions that are not maximally rational. Shifting the focus towards deliberation preceding an action, including partial and prior deliberation, we get a better grasp on agency and autonomy.

Master thesis Ellen Rutten (0457353)  
Supervisor: Dr. Joel Anderson  
autumn-winter 2013-14

## TABLE OF CONTENTS

<b>1. Introduction</b>	<b>3</b>
<i>The debate concerning weakness of will shows there are more actions that we intuitively consider to be rational to some extent, than those that result directly from full, concurrent deliberation.</i>	
<b>2. Method</b>	<b>4</b>
<i>From what we want from a theory of rational action and what we can learn from available views on weakness of will to a liberalized conception of rational action.</i>	
<b>3. What do we expect from a theory of rational action?</b>	<b>6</b>
<i>Some examples of actions that we intuitively like to call rational to some extent and the quest for gaining autonomy.</i>	
<b>4. Some views on weakness of the will</b>	<b>8</b>
<i>Short discussion of the Classical model, Richard Holton, Donald Davidson and Neil Levy on weak-willed action.</i>	
<b>5. Liberalizing the conception of rational action</b>	<b>10</b>
<i>Elaboration of some insights from Holton, Davidson and Levy. Shifting the focus from concurrent better judgment to deliberation so that past and prior deliberation can also be seen as rendering actions rational in some sense.</i>	
<b>6. Meeting the demands</b>	<b>12</b>
<i>Revisiting the examples from section 3 and a look at training rational acting.</i>	
<b>7. Discussion</b>	<b>15</b>
<i>A critical look at the explanatory and pragmatic value of a liberalized conception.</i>	
<b>8. Conclusion</b>	<b>16</b>
<i>A liberalized conception of rational action that focusses on deliberation better meets intuitions and concretely guides the way to gaining autonomy.</i>	
<b>Bibliography</b>	<b>18</b>

## 1. Introduction

Everybody knows at least vaguely what you are talking about when you speak of weakness of the will; choosing one course of action over another, which you believe would be better to take, or perhaps even resolved to take. Whether it's procrastinating something you planned to do, abandoning a resolution that requires self-control or not saying something you think it would be better to say, the phenomenon is familiar to most of us, I believe. Often it may be the case that a weak-willed agent is simply tempted, feels compelled to an action contrary to her better judgment or is overcome or blinded by some other passion. But certainly we are all familiar with cases in which we act very consciously in a way that contradicts our better judgment or previously formed resolution. Sometimes we even make up excuses for acting in a way that could be considered weak-willed in order to feel less guilty and to limit the regret felt afterwards.

There are numerous philosophers that try to explain this common phenomenon, but they all somehow seem to conclude that weak-willed action is an instance of practical irrationality. Their focus is on why and how the agent failed to act (maximally) rational, whereas none of them seem to discuss the possible rationality of such action, except for some who argue that the agent's better judgment or resolution was in fact not maximally rational.<sup>1</sup> The many attempts that have been made to account for the range of weak-willed actions as we experience them, definitely contribute to making the phenomenon more intelligible, but seem to all leave some intuitions unaccounted for. I think the attempts that aim for explaining instances of weakness of the will that we perhaps intuitively would like to regard as rational to some degree, show that the focus in this debate is mainly on the irrational aspects of these cases.

Theorists of rationality tend to speak about the subject in a rather absolute sense; practical reason aims at an optimal course of action. Whether formulated in an 'ought-to-act' or 'ought-to-will' sense or as 'the right cognitive attitudes' in the light of a set of reasons and moral values, it seems that theorists of rationality ascribe as the goal of instrumental rationality figuring out (and willing) the most desirable courses of action. I believe that in order to get a more satisfactory view on agency, we need to revise our model of rational action on a deeper level in order to gain focus on the rational aspects of actions, including instances of weakness of will.

The problem of describing weakness of will in a way that meets our intuitions, including those about the possible rationality of such action, may be a consequence of this all-or-nothing view of practical rationality. Most theorists of rationality seem to describe actions as either rational or irrational, depending on whether the action is in accordance with the agent's concurrent better judgment, all things considered.

Equating rational action with action upon a concurrent, all-things-considered better judgment renders weak-willed actions irrational, regardless of whether they involve deliberation. A person making a complex decision that contemplates some relevant reasons and chooses a course of action different from when he would have considered all reasons is called irrational upon this assumption. I believe most of us would consider this person's action reasonable and would therefore not easily call such action

---

<sup>1</sup> For a short description of philosophers that have followed this line, see Sarah Stroud, "Weakness of Will" in *The Stanford Encyclopedia of Philosophy* Winter 2013 Edition, ed. by Edward N. Zalta. URL = <http://plato.stanford.edu/archives/win2013/entries/weakness-will/#WeaWilPotRat>

plainly irrational. The conclusion that all weak-willed action is irrational directly contradicts some common intuition.

There is another reason to doubt whether the equation of rational action with action upon a concurrent better judgment is a just identification. There seems to be another class of actions that fall outside the scope of rationality on such accounts of rational action, namely that of actions which involve no judgment at the moment of action but are rather the result of a previously formed resolution or trained behavior. In those cases the deliberation about the best course of action took place amply before the action.

The notion that an action only qualifies as rational when it results from a judgment formed after considering all reasons, seems too narrow; many actions that we intuitively consider and commonly refer to as reasonable, fall outside the scope of rational action. In this thesis, I will argue that liberalizing the conception of rational action to also include actions that are the result of past or partial deliberation fits better with our intuitions and will hence serve as a better starting point for any theory of rationality.

## 2. Method

My goal is to defend a conception of rational action that is more concerned with the preceding deliberation than with concurrent, all-things-considered judgment, thus allowing us to regard actions resulting from past or partial deliberation as rational in some sense. In order to do so, I will start by taking a short look at what we expect from a theory of rational action; what should it make intelligible? I will describe some cases which involve action that would be considered irrational on the more stringent conception of rational action, but involve deliberation about reasons and are hence intuitively rational in some sense. I will also look into the question what we wish to gain from a theory of rational action with an eye on improving autonomy<sup>2</sup>.

After roughly establishing the ‘demands’ for a theory of rational action, in section 4 I want to take a brief look at some accounts concerning weak-willed action, which is commonly considered to be a textbook example of a failure of practical rationality. In describing the phenomenon of weakness of the will, theorists of rationality reveal what they consider to be rational action. I’ll start by describing what John Searle calls the Classical Model of rationality. Next I will look at Donald Davidson, who speaks of action upon a judgment that doesn’t include all the arguments, at Richard Holton, who argues that a judgment shift is at cause of weak-willed action and finally at Neil Levy, who describes the phenomenon from a more psychological point of view. In this section I want to show how these views – that either explicitly or implicitly equate rational action loosely with action based upon a concurrent better judgment – are explanatory insufficient.

In section 5 I will elaborate on why I believe the common conception of rational action is too stringent. In a nutshell, this boils down to rejecting the equation of rational action with action upon a concurrent, all-things-considered judgment about the best course of action. I believe that the insights of Holton, Davidson and Levy already carry a seed for broadening the conception of rational action. I want to develop this line of thought on the basis of the examples I provided in section 3, and liberalize the conception of rational action so that these examples fall within the scope of rationality. More

---

<sup>2</sup> I take rational acting to be an execution of autonomy, but don’t mean to equate these concepts.

concretely, I will argue that in order to call an agent rational, deliberation about what course of action is optimal doesn't need to take into account all possible reasons available for the agent in question, or that a clear ranking or resolution of conflicting reasons may be absent. Also, I'll argue that the deliberation about the chosen action doesn't need to take place right before or at the time of action in order for the agent to qualify as acting rationally, but may have taken place in the past – like is the case with an agent that acts upon a resolution without further deliberation at the moment of action, an agent who employs pre-commitment strategies or an agent whose actions follow from deliberately trained habits or behavior.

In short, **I will argue that the deliberation that precedes an action is relevant for the rationality we ascribe to the agent performing that action. This includes (1) partial deliberation and (2) deliberation in the past.** Since the process of deliberation can be more or less thorough and is not always exhaustive, we can also say of actions that they are rational to some corresponding degree. By associating the rationality of an action with the relevant deliberation beforehand, rationality becomes a more flexible matter which can be achieved in various ways. Apart from articulating and weighing of reasons that lead to judgments and intentions, deliberation may also involve reasoning about strategies for committing oneself or training rational behavior.

After articulating this less stringent conception of rational action, I want to investigate in section 6 the strength of this broadened conception as a starting point for a theory of rational action. I aim to show that it performs better when it comes to meeting the demands we have set for a theory of rationality in section 3. First, it makes the concept of rationality intelligible in a way that does justice by our intuitions about rational action, including cases which we consider instances of weakness of will. I will turn again to the examples provided in section 3 and see how the actions described can be regarded rational in the light of this liberalized conception. Second, I will look at the tools for enhancing rational acting that we can gain from a theory of rationality that employs a broadened conception of rational action; making the various ways in which we can act rationally intelligible paves the way towards articulating strategies that stimulate rational decision making.

To make this more tactile, I will discuss some articles concerned with psychological research that I think are relevant, because they are concerned with the process of deliberation involved in the decision making process. This will include Daniel Kahneman's observations about thinking fast and slow, and the idea of an effort-reduction framework as sketched by Shah and Oppenheimer. This kind of research is the basis on which some scientists and philosophers are trying to articulate ways to increase and improve rational acting, much like I suggest should be possible when conceiving rational action as action upon deliberation. I will say some things about Eliezer Yudkowsky, and refer to the community blog Less Wrong that he founded and the subsequently formed Center for Applied Rationality that aim at teaching strategies for acting more rationally. Apart from these initiatives, I'll briefly point at other ways to improve rational acting that seem to be implicitly build upon the conception of rational action articulated in this thesis.

In conclusion, I hope I will convince the reader that a liberalized conception of rational action that focusses on deliberation preceding an action rather than on concurrent judgments, serves as a better starting point for any theory that aims to make rational action intelligible in a way that meets our intuitions and paves the way towards gaining autonomy, because it emphasizes the rational aspects of actions that are not maximally rational.

### 3. What do we expect from a theory of rational action?

Since I claim that theories of rational action proposed thus far fail to some degree in properly explaining our intuitions and since I am proposing to change our conception of rational action to conduct towards a better theory, I think it is important to take a closer look at what constitutes a ‘good’ theory of rational action. What should such a theory make intelligible? Which intuitions should it meet, which phenomena should it explain? And should it provide us with insights to help us become more rational?

#### Explaining our intuitions

First of all, it is clear that we expect a theory of rational action to make the phenomenon of acting rationally intelligible. Here it is immediately clear that we use the term ‘rational’ both to refer to an action itself and to refer to an agent choosing and performing an action. I think this twofold employment of the term ‘rational’ carries with it the risk to mix up the agent performing an action with the action itself. Since we commonly regard rationality as a human capacity, I think it makes sense to say that in the first place, an agent is rational in acting in some way and that we call an act rational as a result. A theory of rational action should thus tell us when we ascribe rationality to an agent choosing a certain course of action.

Plenty of accounts have been offered to settle this question about ascribing rationality, but there are numerous examples of actions that seem to be hard to capture in a satisfying way. These examples include, as I already mentioned in the introduction, cases of weakness of the will. Many theories seem to have difficulties dealing with those examples in a way that matches all our intuitions, or alternatively explains how and why our intuitions about those cases are mistaken. It is often said that a weak-willed act is the result of a failure of practical rationality. On these accounts, the weak-willed agent is described as acting irrationally. This is not in line with our intuition that there is something rational about some weak-willed actions; thinkingly and knowingly acting in a way we know we (would) judge inferior to some other course of action is not a matter of blind irrationality, we would say. In some cases, we do deliberate about our weak-willed action; we bear in mind our better judgment about the best course of action or we make up excuses why our better judgment is perhaps faulty. Furthermore it can happen that an agent has difficulties ascribing the proper weight to her reasons or doesn’t resolve their evaluation before acting and instead follows her intuition.

There are also examples of agents acting in a way that we would intuitively consider to be rational, where the agent does not bear in mind any reasons for that action and doesn’t deliberate about the best course of action. Take an agent that acts on a previously formed resolution and purposely refrains from holding any judgment about the preferable course of action at the moment of action. I think most of us will agree that it can be very rational to act upon a resolution without further deliberation about whether the resolution matches concurrent desires.

In order to check how well a theory performs at explaining our intuitions about the rationality of actions, I think it will be useful to make use of some examples reflecting actions of the sorts mentioned above. I designed the following examples to contain actions that involve deliberation at the time of action that may be regarded as incomplete (Nikki) or corrupted by temptation (Tim), as well as deliberation in the past that led to a resolution (Tamara) or a self-regulatory strategy (Simon). This

way, these examples reflect the various ways of deliberating about action that I think are commonly not reflected sufficiently within philosophical theorization.

1. Nikki has always wanted a dish washer. She really hates doing the dishes after dinner and usually leaves them for days. This often results in her cat breaking something at night by knocking it off the counter. She has little money and knows that all things considered, the best thing to do would be to simply do the dishes after dinner, thus saving money, water and solving the cat problem. Nevertheless, on a day when she found broken plates on the kitchen floor once again, she goes out and buys herself a dishwasher, realizing that she will have to borrow money from her parents to pay the bills this month.
2. Tim is a writer who has resolved to drink less for health reasons and allows himself a maximum of two glasses of wine per day. On a Thursday, he has a somewhat disappointing health check-up at the doctor. That night he drinks four glasses of wine at dinner, even though he was planning to do some writing that evening and knowing that he will probably feel some regret the next day.
3. Tamara loves to watch television at night, but as a result, she usually goes to bed later than planned and never gets to reading, which she regrets. One day, she decides that on weekdays, she will switch off the television at 10 pm, after her favorite show, so she has some time to read before she goes to bed, in time. This soon becomes a habit; at the ending credits, she automatically gets up from the couch and switches off the television.
4. Simon loves to eat some chocolate with his coffee after dinner. Usually he cannot control himself though and eats more than he judges healthy and as a result is becoming a little overweight. In order to prevent this from happening, he made an agreement with his wife: she always handles the chocolate in the house and gives him one piece of chocolate with his coffee. One day, the neighbor gave Simon two bars of chocolate as a thankyou for helping him out. Simon immediately gives the chocolate to his wife.

### **Pragmatic value**

According to Searle, a common expectation of a theory of rationality is that it provides an algorithm for rational decision making. He argues against this expectation and says that a theory of rationality cannot tell you how to make the right decisions, the same way a theory of truth doesn't provide an algorithm for discovering true propositions.<sup>3</sup> I believe this analogy is quite useful, but I also think that Searle draws too radical a conclusion from the boundaries he sees for theories of truth and rationality. Indeed, a theory of truth won't guide towards the discovery of true propositions, it nevertheless shows how to settle the question whether we should call a certain proposition true or not. I think a theory of rationality should in the same sense be able to guide us towards deciding whether we find an action rational. If a theory of rationality makes intelligible when we ascribe rationality, it logically follows that we can use such a theory to answer questions about the rationality of particular actions. In this sense I believe that, although a theory of rational action cannot simply decide what would be a rational course of action regardless of the agent's reasons or beliefs, desires, values etc., it should nevertheless help us understand or even anticipate on which grounds a certain action should be considered rational upon that theory.

---

<sup>3</sup> John R. Searle, *Rationality in Action* (Cambridge: The MIT Press, 2001), XV.

Gaining insight in ascribing rationality to an acting agent and hence to certain actions, is perhaps the most relevant pre-condition for when it comes to figuring out strategies for acting more rationally. Once we have settled the question about under which conditions we can regard an action to be rational, we can start to find strategies for fulfilling these conditions.

In conclusion, I think the main demand for a theory of rationality is to make intelligible when we ascribe rationality to an acting agent in a way that fits with or explains our widely shared intuitions. This includes particularly making good sense of the examples I described. The better a theory succeeds at providing a conception of rational action that covers our intuitions, the more insight it provides in the conditions in which we should call an action rational. I think that from this insight follows naturally a better grip on ourselves as rational agents; understanding at least which actions are rational is a logical first step in working towards creating tools and guidelines for acting more rational.

## 4. Some views on weakness of the will

In this section I want to examine a few ways of looking at rational action by examining how they explain the phenomenon of weakness of will. I will start by following Searle in his description of the Classical Model to see how weakness of the will is classically accounted for and what rational action is on this model. Next, I will briefly sketch the moves made by Holton, Davidson and Levy to try and make better sense of the phenomenon, bearing in mind the examples from the previous section. I will shortly review the discussed approaches, thus setting the stage for a revised conception of rational action in section 5.

### The Classical Model

According to Searle's description of the Classical Model of rationality<sup>4</sup>, rationality is an instrumental and separate cognitive faculty concerned with following rules of rationality – whatever they may be. Rational action is traditionally considered to be the causal result of beliefs and desires. Akrasia, understood as acting against one's better judgment (all things considered), is strictly impossible on this view. For when the better, all-things-considered judgment is present, the corresponding action will follow necessarily. All cases in which we speak of weakness of will are, according to this model, cases in which something is wrong with the psychological antecedents of the action; the better judgment is not properly formed or present. This very straightforward way of describing rational action results in the conclusion that weak-willed action cannot be considered rational.

Perhaps many presumed cases of weak-willed action can be explained this way – the agent didn't think through her action very well or is overcome by passion – but it doesn't seem to do justice by all the aspects concerning the cases of Nikki and Tim, who seem to act contrary to a better judgment in a way that involves a conscious process of weighing reasons, even contemplating the previously formed better judgment. They are not blinded by passion or simply overcome by passion. Nothing much seems wrong with the psychological antecedents of the action; they both appear psychologically sane to us. Even though they are perhaps not maximally rationally, there is something reasonable about their actions that is not captured by the classical model. The cases of Tamara and Simon cannot be

---

<sup>4</sup> Ibid., 5-12.

considered cases of rational action on the classical model; the actions are not the direct result of a judgment, let alone a better judgment.

### **Richard Holton and judgment shift**

Richard Holton aims to account for the phenomenon of weakness of will by first distinguishing it from akrasia. Akrasia is acting against a better judgment, but he says that in cases of weakness of will there is no better judgment involved.<sup>5</sup> According to Holton, the weak-willed agent abandons a resolution when facing temptation. Weakness of will, he says, is the unreasonable revision of a resolution that was made in order to defeat a contrary inclination. This revision is unreasonable when the resolution is revised under the pressure of exactly this inclination. So what is key to Holton's view is that a weak-willed agent has a certain resolution and revises this when facing the temptation she was trying to defeat. Typically, the agent will regret this afterwards; her judgment shifted in the face of temptation and shifts back afterwards. Later on, Holton further develops this line of thought and argues that it can actually be rational to not reconsider a resolution<sup>6</sup>; this way the action will match the better judgment upon which the agent previously based her resolution, and not a potentially shifted judgment that may follow from reconsideration.

Holton describes an important intuition when it comes to acting irrationally; it often involves a temporary judgment shift in the face of temptation, characterized by regret afterwards. But what about Nikki? She may have acted upon a shifted judgment, and maybe has a hard time handling her bills for a few months, but nevertheless doesn't regret her purchase. We wouldn't want to call her purchase completely irrational as a result, but according to Holton, her action nevertheless is irrational. In the case of Tim; his over-all judgment about drinking more than two glasses hasn't shifted, he knows that all things considered it would be better not to, even while drinking. It may nevertheless be too one-sided to say that he is acting irrationally. He may feel no or little regret afterwards and simply continue with his two-glass-policy the next day. It may be reasonable for Tim to make this exception; he may be thinking to himself that he had a rough day and he is only human, so an exception is acceptable. The cases of Tamara and Simon are not that puzzling from Holton's point of view: since he regards non-reconsideration of a resolution to be rational, it seems reasonable to call their actions rational on his account. For Tamara this is self-evident, since she unambiguously acts upon a resolution, while the case of Simon, who employs a pre-commitment strategy, isn't a straightforward example of rational non-reconsideration, but seems to touch the same idea.

### **Donald Davidson and unconditional judgments**

According to Davidson<sup>7</sup>, weakness of will should be seen as incontinent action. The paradox of incontinent action arises when we both assume that incontinent action exists, and that it is possible for an agent to judge one possible course of action better, all things considered, but still voluntarily perform another action. Davidson argues that an agent acts incontinently when she acts not upon such an all-things-considered judgment, but rather on a judgment that doesn't include all the arguments. In

---

<sup>5</sup> He describes that intentions, not better judgments play a crucial role in weak-willed action: "I shall develop the idea that the central cases of weakness of will are best characterized not as cases in which people act against their better judgment, but as cases in which they fail to act on their intentions" (Richard Holton, *Willing, Wanting, Waiting* [Oxford: Clarendon Press, 2009], 70).

<sup>6</sup> *Ibid*, 140.

<sup>7</sup> Donald Davidson, "How Is Weakness of the Will Possible?" in *Essays on actions and events*, 21-42 (Oxford: Clarendon Press, 2001).

other words: weakness of will is a consequence of not bearing all the relevant reasons in mind. This is a failure of practical rationality.

The problem with this view is parallel to the problem with Holton's view: it renders all weak-willed or incontinent action irrational, since it still goes against the agent's all things considered judgment. Davidson does explain the intuition that an agent can act weak-willed while bearing in mind reasons; the agent simply doesn't consider *all* the relevant reasons. The actions of Nikki and Tim should thus be labelled 'irrational' also on his view.

### Neil Levy

Leaning heavily on cognitive and social psychological research, Neil Levy argues that the concept of weakness of will is one we should abandon. He basically argues that cases of 'weakness of the will' are cases of a broader phenomenon, namely an agent switching from one sort of cognitive processing to another. To follow his reasoning, it is necessary to understand what these two systems of cognitive processes should be seen like. In his article 'Resisting 'Weakness of the Will'' he describes those as follows:

*"System 1[...] consists of a set of mechanisms that respond automatically to stimuli, without the need for oversight from consciousness. System 1 processes are fast, ballistic and undemanding of cognitive resources [and] operate in parallel. System 2 has the opposite profile: it consists of mechanisms that are slow, operate serially rather than in parallel, and are demanding of cognitive resources. System 2 processes are rule-governed and conscious."*<sup>8</sup>

According to Levy, cases of weakness of will are cases where the agent performs a system 1 action that differs from the action judged better by system 2. I think Levy makes a very interesting and fruitful connection between dual-process theory coming from psychological research and the philosophical debate concerning weakness will. However, Levy's proposed identification of weak-willed action with action resulting from system 1 cognitive processing seems to fail when we regard the examples of Nikki and Tim: we probably want to say that their actions are weak-willed to some extent, but nevertheless they seem to result from system 2 processes. The actions of Tamara and Simon actually seem to be the result of system 1 processes.

## 5. Liberalizing the conception of rational action

We have seen in the previous section that theories of rational action and accounts of weakness of the will that employ a conception of rational action as action upon a concurrent, all-things-considered judgment, don't do very well in meeting all our intuitions about rational action. It seems to me that the traditional conception of rational action is too stringent and that as a result many actions that we intuitively consider and commonly refer to as rational, whether genuinely or to some degree, fall outside the scope of rational action theory. After analyzing the examples I use in this thesis it becomes clear that the traditional conception is too stringent in two aspects. First, the demand for all things having been considered or all reasons being taken into account seems too high. Second, it focusses too

---

<sup>8</sup> Neil Levy, "Resisting 'Weakness of Will'," *Philosophy and Phenomenological Research* LXXXII-1 (2011): 134-55, 145.

exclusively on the deliberation and judgment at the time of or right before the action. I will deal with the two senses in which I believe the traditional conception needs to be liberalized separately.

### **Partial deliberation**

With the help of the examples of Nikki and Tim we have seen that it may happen that an agent decides towards an action after considering some, but not all reasons relevant to the action. It goes against our intuition to say that all these actions are plainly irrational. Though perhaps it is right to say that the cases of Nikki and Tim are cases of weak-willed action<sup>9</sup> – after all, there was a better course of action possible that a very strong-willed person may have chosen to take – they did not blindly fall for temptation, but deliberated about their actions. If Nikki hadn't thought about any reasons for buying a dishwasher but did so impulsively, we would probably have considered her action to be simply irrational. Likewise, if Tim hadn't thought about a reasonable excuse for himself to make an exception on his policy, but compulsively would have drunk a bottle of wine, we also most likely wouldn't have regarded that action as very rational. This seems to imply that the deliberation that took place is what makes us intuitively consider their actions to be rational to some extent. The fact that they didn't exhaustively consider all possible reasons – Nikki 'conveniently' left out her own laziness and Tim didn't give a lot of thought to the conditions under which he would find exceptions reasonable – and thus didn't come to an all-things-considered judgment about the best course of action, may perhaps lead us to believe their actions aren't maximally rational, yet it doesn't convince us they are best labeled irrational either.

More generally speaking, I believe it is very often the case that agents act upon a subset of reasons and do not consciously contemplate all possible reasons. I want to call this sort of thinking preceding an action *partial deliberation* since only part of the reasons is involved in the deliberation or the reasons are not considered to a maximum extent. This observation is nothing new: Davidson refers to this phenomenon in explaining incontinent action. The judgment shift that Holton describes may also be seen in this light; direct temptation gives rise to a new or stronger reason and disturbs the previously taken stock. They describe this kind of behavior, but don't go as far as to say that we are maybe better off calling this behavior rational to some extent. The focus remains on where and how the agent failed in forming or acting upon a better judgment. I think it is important to articulate not only the failure, but also the successes of rationality concerning such action, so that we get a more complete picture of rational action.

### **Prior deliberation**

The cases of Tamara and Simon show that there are situations in which an agent acts in a way that we would intuitively label rational, whereas no deliberation about reasons occurs or no judgment whatsoever is present at the time of action. The most enlightening case of such rational action is perhaps the mindless following through with a previously formed intention or resolution. The weighing of reasons in order to arrive at a resolution may lie far in the past relative to the unthinking execution of an action upon this resolution, yet we would surely like to call acting upon a well formed resolution rational. Instead of forming a genuine resolution, an agent can also deliberate about how to

---

<sup>9</sup> The case of Nikki may be considered to not be an instance of weakness of will, since she didn't explicitly conclude to an all-things-considered judgment and she doesn't feel regret. However, this labeling isn't crucial for the argument.

train behavior, as we have seen with Simon. I think we would like to call actions that are the result of trained behavior or previously thought-through automatisms rational as well.

The weighing of reasons far prior to an action, like in forming an intention or a resolution as well as the reasoning involved when it comes to training behavior and consciously shaping habits, I want to call *prior deliberation*. The suggestion to call actions that are in line with past deliberation is also nothing new; Holton elaborates on the phenomenon of resolution making and actually calls the following through with resolutions without further deliberation ‘rational non-reconsideration’<sup>10</sup>. He thus already leaves room for rationality to be predicated of unthinking action. I think this is a very important insight in rational acting that should also be incorporated in a new conception of rational action.

### **New conception**

As stated above, I believe it is necessary that we liberalize our conception of rational action as to incorporate actions that are reasonable to some extent, like we have seen in the four examples in section 3. I believe we should let go of the traditional stringent conception of rational action that rational action involves exhaustive deliberation to form a better judgment at the moment of action. Though this latter condition of concurrency is perhaps often not specifically articulated, I believe that explicitly rejecting it is important, in line with Holton’s explication of the phenomenon of rational non-reconsideration. It think that also we need to let go of the too absolutist demand implied in ‘all-things-considered’; that in order for an acting agent to qualify as rational, she should consider exhaustively all reasons or bear in mind a judgment that ultimately serves best her desires and ends.

In other words, I think a conception of rational action should allow past or partial deliberation to render an acting agent, and hence her action, rational to some degree. I propose to correlate the rationality of an action with the deliberation that preceded the action. Similarly, an agent should be considered rational insofar she deliberated about her action. This deliberation may include seeking reasons, ascribing them the proper weight, forming judgments, intentions and resolutions, training behavior and habits, deploying pre-commitment strategies and so forth.

## **6. Meeting the demands**

In the section above I have argued that the traditional conception of rational action falls short in two ways. I have proposed that we let go of the too stringent demands that an action be based upon both concurrent and all-things-considered better judgment in order to qualify as rational. Instead, we should work towards a theory that describes how past and partial deliberation can also grant the status rational to an action, possibly in degrees.

After explicating some grounds for a liberalized conception, I now want to examine what we can gain from this new perspective; does it perform better when it comes to meeting the demands we have set for a theory of rational action in section 3? Does it fit better with our intuitions about rational action? If so, we should be able to explain why the actions from the examples appear to us as rational to some

---

<sup>10</sup> Holton, *Willing, Wanting, Waiting*, 140.

degree. Also, from this insight we should be able to make sense of the processes involved in acting rationally, guiding towards strategies to improve these processes.

### **Explaining intuitions**

When it comes to meeting our intuitions considering rational action with regard to the examples I provided earlier, it will come as no surprise that the proposed liberalized conception is able to capture those, since they were the starting point of my argument towards this liberalization. Nevertheless, I think it is useful to complete this argument by showing how these intuitions are reflected in this wider conception of rational action.

Let us start by looking at Nikki and Tim. Both base their action upon reasons they deliberate about in the present – Nikki muses about the benefits of a dishwasher and Tim uses his disappointing doctor visit as a reason to make an exception on his otherwise reasonable policy. This partial deliberation is enough ground to call their action rational in the broadened sense of rational action. We may need to look further into the exact degree to which these actions are rightfully called rational, nevertheless they sufficiently qualify to be considered rational actions.

The actions of Tamara and Simon don't involve deliberation about reasons at the time of action but are rendered rational due to the past deliberation involved. Tamara thought about her reasons for watching television and reasons to want to limit that, weighed them, formed a resolution and a concrete plan to make following through with what she judged the best course of action feasible. Simon likewise deliberated about his reasons concerning eating chocolate. The policy that seemed to reflect the balance between his reasons pro and con served as a ground for making the agreement with his wife.

More generally speaking, we see that associating the rationality of an action with the deliberation concerning that action, leaves room for regarding more actions rational than most theories do. It provides grounds for calling actions rational that are the result of weighing *some reasons, but not all*, as well as actions that result from considering all relevant reasons, where the agent *didn't complete the weighing process* – which may be due to time limitation, inability to properly rank the reasons or something else. Furthermore, unthinking action resulting from deliberately *trained behavior* or habits or the unthinkingly following through with deliberately made *intentions or resolutions* are also rational. The new conception of rational action thus seems to be more flexible by incorporating these actions within the scope of rationality. Actions that we intuitively regard to be at least somewhat rational, because they involve some form of deliberation.

### **Acting rationally**

The second demand for a theory of rational action is that it makes rationality intelligible in a way that may guide conduct toward increasing rational decision making. By explicating the various ways in which we can act rationally, the preconditions for rational action have become clearer. This is the first logical step towards figuring out strategies to improve our rational agency. Understanding better when and how we act rationally guides towards creating the appropriate preconditions for rational action.

Understanding that the rationality of an action is something we intuitively equate with the deliberation that preceded it, we can start to make sense of this process and its functions. If deliberation is what makes an action rational, it is important to understand this phenomenon better. In doing this, it is most useful to look at psychological research concerning deliberative thinking. Next, we can start to

investigate concretely the ways in which we can promote deliberation and how we can make good use of it.

A very insightful book on the way people make decisions, with or without deliberation, is *Thinking, Fast and Slow* by Daniel Kahneman. He describes the two ways in which humans process information and come to decisions: in a fast and unconscious manner or via slow and conscious deliberation. His observations correspond to the presumed two psychological systems responsible for decision making, as also described by Neil Levy. He describes research showing that we indeed sometimes deliberate about our actions, and sometimes we act upon something like an auto-pilot. Levy even writes that most of our actions are the result of system 1<sup>11</sup>. There is a lot of contemporary psychological research about how we make decisions about our course of action. The heuristics and biases program aims at describing ‘shortcuts’ and other mechanisms that allow an agent to make decisions faster and more easily. Some interesting research concerning heuristics is described by Anuj Shah and Daniel Oppenheimer, who promote an effort-reduction framework to help understand how, when and why people employ heuristics.

Using these insights, Eliezer Yudkowsky<sup>12</sup> started a community blog called Less Wrong<sup>13</sup> devoted to the topic of rationality, with quite some focus on improving the process of rational decision making. For example, heuristics are explained so that the reader may gain insight in his own psychological mechanism and hence increase the quality of deliberation. This blog gave birth to the Center for Applied Rationality<sup>14</sup>, a community promoting rationality in daily life via for example workshops. On their website you can find a checklist that should help *develop rational habits*.

These initiatives, sparked by psychological research, show that it is widely believed that agents can improve their rational decision making by means of deliberation. Their common aim is not to help make a perfectly rational decision, but to gain insight in one’s reasons, weigh them critically and successfully develop habits to act in accordance with evaluative assessments.

It may be questionable whether an agent can increase how much or how often she deliberates, but initiatives like Less Wrong and the Center for Applied Rationality at least aim for *increasing the quality of deliberation* to increasing rational decision making and hence rational acting.

Understanding that we sometimes fail to deliberate about our actions, which could be explained as the inevitability for acting upon system 1 at times – what Isabelle Bauer and Roy Baumeister describe as the finiteness of self-regulatory strength<sup>15</sup> – it becomes evident that training our habitual system 1 may contribute to acting in a way that system 2 decided earlier would be best. Understanding our biases and heuristics as well as training our behavior are ways of procuring the rationality of actions upon system 1.

Besides increasing the quality of deliberation or training habits or behavior, an agent can also deploy her environment for gaining autonomy. Knowing that our deliberating system 2 is confined, we could

---

<sup>11</sup> “[T]he overwhelming majority of our actions are produced by system 1 which takes care of our basic survival needs and much more besides.” Levy, “Resisting ‘Weakness of Will’,” 145.

<sup>12</sup> <http://yudkowsky.net/>

<sup>13</sup> <http://lesswrong.com/>

<sup>14</sup> <http://rationality.org/>

<sup>15</sup> Isabelle M. Bauer and Roy F. Baumeister, “Self-Regulatory Strength,” In *Handbook of Self-regulation: Research, Theory, and Applications*, edited by Kathleen D. Vohs and Roy F. Baumeister, 64-82 (New York: The Guilford Press, 2011).

in advance take measures to restrict our options at the time of action. When for example a temptation is absent, an agent can deliberately employ *pre-commitment strategies* so that when temptation presents itself, it is harder or even impossible for the agent to act contrary to her previously formed judgment about the desirable course of action. A good example of this is Simon who made an arrangement with his wife about the chocolate. The risk of system 1 taking over control after dinner so that he would eat too much chocolate is eliminated beforehand, when he was using system 2 in deliberating about ways to eat less chocolate. But there are plenty of ways in which an agent can bind herself, or as Joseph Heath and Joel Anderson put it<sup>16</sup>, scaffold their intended rational actions. Think about for example publicly sharing a resolution to induce shame when one would fail to follow through, thus creating an extra reason. Another method is making checklists to force oneself to stay aware of intended goals. Or using a punish or reward system.

Understanding the crucial factor of deliberation for rational acting, allows us to investigate ways to improve and make better use of this process. Whether it is by gaining insight in heuristics and biases that may corrupt our deliberation or by using deliberation to train rational habits or scaffold rationality. We thus see that a theory of rationality that employs a liberalized conception of rational action is a fruitful starting point for finding and creating tools and strategies that can help us act more rationally.

## 7. Discussion

Changing the conception of rational action to shift focus more towards deliberation will probably not go without meeting new problems, or old problems in a different form. There are many views possible when it comes to describing rational action and one may question what exactly we gain and what we lose from changing the focus towards deliberation compared to available theories. We need to look critically at both the explanatory and the pragmatic value of this conception.

I proposed to liberalize the conception of rationality in order to better capture some presumed intuitions. By focusing on the deliberation involved, the identifying scope of rational action became broader, but what if it now also identifies rationality where we would intuitively not? By letting go of the demand of concurrency of judgment, we risk having to ascribe rationality to an agent that unthinkingly acts upon an intention or resolution there is now good reasons to abandon, or an agent acting as a result of behavioral training that does no longer suit the agent's regular setting or is applied in an inappropriate context. If Tamara for example has some vacation days and she is watching television with a friend, automatically switching it off after her favorite show is suddenly a lot less rational. Perhaps her friend's favorite show even starts after Tamara's favorite show. There are reasons to say that the habit is inappropriate in this context. Suppose she receives a new schedule for work; she starts an hour later every morning and can go to bed an hour later. After her favorite show, there is another show on television she really enjoys. This may be a good reason to say her trained behavior of switching off the television after her favorite show does no longer suit her daily pattern very well and it may be rational to change it. Perhaps even clearer is the story of Patrick who quit smoking when he turned 25 and resolved to buy a motorbike from the money he saved this way when he turns 35. Now that he is 35, he has a family and some debts. He has a lot of good reasons to not buy the motorbike and spend the money more responsibly.

---

<sup>16</sup> Joseph Heath and Joel Anderson, "Procrastination and the Extended Will," In *The Thief of Time: Philosophical Essays on Procrastination*, edited by Chrisoula Andreou and Mark D. White, 233-52 (Manuscript, 2009).

These examples show that by letting go of the concurrency demand, we may be ascribing rationality in more cases than our intuition tell us is right. Certainly there are cases in which it is rational to not reconsider an intention or resolution, as Holton already argues, but there are also cases in which reconsideration is actually the more rational thing to do. We need to prevent having to ascribe rationality to agents blindly following through with an intention or resolution or deliberately trained behavior in a context where we would generally consider it more rational to reconsider. We expect an agent to respond to relevant changes in herself or her environment. To determine when it is more rational to refrain from deliberating about an intention, resolution or habitual behavior and when it is more rational to deliberate, for instance about newly presented reasons, we may take a look at Holton, who discussed this matter in his book *Willing, Wanting, Waiting*. He describes several conditions<sup>17</sup> under which it is or is not rational to reconsider, in an attempt to better capture the phenomenon of rational (non-)reconsideration. I think that shifting our focus towards deliberation means we have to thoroughly investigate when and why we want to call actions resulting from prior deliberation rational in some sense.

I also argued that focusing on deliberation helps us to find starting points for gaining autonomy, but one may ponder whether letting go of the demand of all things having to be considered doesn't cause a decrease in motivation for acting more rationally. If we ascribe rationality to anyone deliberating, what do we set as our goal? Doesn't for example the Classical Model, that labels weak-willed action irrational regardless of the deliberation involved, serve as a better starting point for gaining autonomy by pointing at the failures, where the room for improvement can be found? Are we 'lowering the bar' by saying that deliberative action is rational in some sense? I think it is indeed important to not solely focus on the rational aspects of actions, since the irrational aspects require attention if we want to improve our autonomy. When we associate deliberation with rationality, thus ascribing rationality to a lot more actions, we should not lose sight of the deviation between an action that is rational in some sense and a maximally rational action. I believe it is very important that we complement this liberalized conception of rational action with the articulation of the irrational aspects of our actions.

## 8. Conclusion

I have argued that a theory of rationality that employs a conception of rational action as action upon a concurrent, all-things-considered better judgment doesn't do justice by our intuitions. This struggle to explain widely shared intuitions is reflected in the ongoing debate around weakness of the will. In trying to account for weak-willed action, some philosophers already lay bare the fact that people sometimes deliberate in a suboptimal way, purposely refrain from deliberating at the time of action or deliberate about ways to pre-commit oneself or scaffold rationality. I have argued to extrapolate these observations to a new conception of rational action. The examples that I designed to reflect actions that would theoretically be called irrational, but we commonly intuit as rational in some sense, showed that the rationality we ascribe to certain actions is correlated to the deliberation that preceded that action.

Correlating the rationality of an action with the relevant deliberation beforehand, leads to a wider scope of rational action; it now includes actions that follow from what I call partial and past deliberation. This new conception captures the rational aspects of the actions from the examples and

---

<sup>17</sup> Holton, *Willing, Wanting, Waiting*, 160.

thus helps us gain insight in how to act rationally. Instead of focusing on forming all-things-considered judgments at the time of action, this approach guides to the process of deliberation and the various ways we can make use of it. The initiatives for improving applied rationality by gaining insight in the process of deliberation as well as the strategies to employ deliberation strategically to ensure the desired action will be taken at a non-deliberative moment, show the success of this focus on deliberation.

A theory of rational action that focusses more on this process of deliberation and its various applications explains more of our intuitions, provides grip on autonomy and seems to fit well with psychological research concerning decision making. I have explored this conceptual liberalization only briefly and in an abstract sense, so naturally a lot more of philosophical investigation may be in place to get a better understanding of the consequences. However, I hope that my argument will convince philosophers to investigate moving their conception of rational action towards the line that psychologists seem to be following in practice.

## BIBLIOGRAPHY

- Ainslie, George. "Précis of *Breakdown of Will*." *Behavioral and brain sciences* 28 (2005): 635-73.
- Bauer, Isabelle M. and Roy F. Baumeister. "Self-Regulatory Strength." In *Handbook of Self-regulation: Research, Theory, and Applications*, edited by Kathleen D. Vohs and Roy F. Baumeister, 64-82. New York: The Guilford Press, 2011.
- Davidson, Donald. "How is weakness of the will possible?" In *Essays on actions and events*, 21-42. Second edition. Oxford: Clarendon Press, 2001.
- Evans, Jonathan St. B. T. "In two minds: dual-process accounts of reasoning." *TRENDS in Cognitive Sciences* 7:10 (2003): 454-9.
- Heath, Joseph and Joel Anderson. "Procrastination and the Extended Will." In *The Thief of Time: Philosophical Essays on Procrastination*, edited by Chrisoula Andreou and Mark D. White, 233-52. Manuscript, 2009.
- Holton, Richard. *Willing, Wanting, Waiting*. Oxford: Clarendon Press, 2009.
- Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.
- Levy, Neil. "Resisting 'Weakness of Will'." *Philosophy and Phenomenological Research* LXXXII-1 (2011): 134-55.
- Muehlhauser, Luke ('lukeprog'). "How You Make Judgments: The Elephant and its Rider." Accessed February 28, 2014. URL = [http://lesswrong.com/lw/531/how\\_you\\_make\\_judgments\\_the\\_elephant\\_and\\_its\\_rider/](http://lesswrong.com/lw/531/how_you_make_judgments_the_elephant_and_its_rider/)
- Searle, John R. *Rationality in Action*. Cambridge: The MIT Press, 2001.
- Shah, Anuj K. and Daniel M. Oppenheimer. "Heuristics Made Easy: An Effort-Reduction Framework." *Psychological Bulletin* 134:2 (2008): 207-22.
- Stroud, Sarah. "Weakness of Will." *The Stanford Encyclopedia of Philosophy* Winter 2013 Edition, edited by Edward N. Zalta. URL = <http://plato.stanford.edu/archives/win2013/entries/weakness-will/>