What is Rational Action?

An Argument for a Particular Version of The Humean View

by

Bendik Berntsen-Øybø

Master's thesis in philosophy

Supervisor: Professor Caj Strandberg

Master's Program in Philosophy

University of Oslo

Department of Philosophy, Classics, History of Art and Ideas

Fall 2019

What is Rational Action?

An Argument for a Particular

Version of The Humean View

2019

What is Rational Action? - An Argument for a Particular Version of The Humean View

Bendik Berntsen-Øybø

http://www.duo.uio.no/

Abstract

This thesis is a contribution to the endeavor of uncovering what the concept of rational action is. In it, I argue that it seems reasonable to pursue an understanding of rational action based on The Humean View, that is, based on the presumption that there is only one requirement an action needs to meet in order to be a rational action, namely an instrumental requirement. After having done this, I consider how we should understand this requirement. By drawing on the literature on the topic, I identify nine different questions which it seems sensible to ask with regard to how this requirement is to be interpreted. I argue for a particular answer to each of these questions, and based on them, I formulate an understanding of the instrumental requirement and propose that we understand rational action as an action that meets the instrumental requirement under this understanding. I conclude that it seems reasonable to think that a rational action is an action which the evidence of the actor suggests will best help fulfill the totality of all of the contemporary intrinsic desires of the actor, and the actor intends to try to achieve this goal by performing the action.

Acknowledgements

My supervisor for this project has been Professor Caj Strandberg. The discussions we have had have often been eyeopening, and the insights he has shared have been treasured. His feedback on my drafts was essential to keep me on the right track. For his guidance and support throughout the process, I am very grateful.

The support I have received from my fellow graduate student Maria Luz Crebay, has been greatly valued. It has been a pleasure discussing the topic of my thesis with her, and on receive her thoughts and feedback to what I have been writing.

Last but not least, I would like to thank Catherine, my wife, who has been immensely patient and supportive during these many months of writing, and who has lifted my spirits every day.

Table of Contents

1 - Introduction

1.1 - Concerning the discussion of the thesis

The goal of this thesis is to establish the reasonableness of a particular account of what rational action is as a concept. The topic of what rational action is has been frequently discussed during recent years. Philosophers have given surprisingly many different answers to the question. This illustrates how difficult it is to figure out how to best understand the concept of rational action. What this is likely to be partially attributable to is the fact that many different questions can be asked with regard to the nature of practical rationality, and philosophers have found promising many different answers to these questions. Because of this, there are very many lines of division. One line of division is based on how many types of requirements of rationality one thinks there is. With regard to this, some philosophers hold The Kantian View, which says there are multiple, while others hold The Humean View, which says that there only is one, namely an instrumental requirement. The idea of the instrumental requirement is loosely the idea that one is rationally required to take the means to one's goal or goals. However, what fulfills this type of requirement is not a given. For example, one can hold different views with regard to what determines this type of goal. One can hold the view that this goal is only determined by what the actor intends to achieve. Alternatively, one can hold the views that it, in addition, is determined by what the actor desires, what is good for the actor, or what is independently valuable. I will come to argue for the view that it is determined by the desires of the actor in addition to her intentions. We can differentiate between many types of desires, however, and exactly what type of desires determine the relevant goal of the actor is also a matter of dispute. There are many different available views with regard to what criteria a desire must meet in order to be this type of relevant desire. One can, for example, hold the view that desires one will have in the future are relevant desires or the view that only desires one has when one is performing the action in question determines the rationally relevant goal for that action. Another central line of division is whether the instrumental requirement can be fulfilled by actively changing the relevant goal one has. Here, the discussion stands between those that think that it can and those that think the

requirement only can be fulfilled by taking the prescribed means. These and even more lines of division will be discussed throughout this thesis.

There are at least three important reasons why understanding what rational action is important. Firstly, it is commonly thought that *what is rational for people to do* has at least some degree of correlation with *how they tend to behave*. Given such a correlation, knowledge about what is rational for individuals to do can be used to help predict human behavior. In order to understand how strong this correlation is, why this correlation exists, and why it is of the strength it is, it is crucial to understand what rational action is.

Secondly, phrases such as *you are about to act irrationally* and *I believe this action would be rational for you to perform*, if believed by the person who is being talked to, tends to be very powerful in terms of making the person reconsider what action to perform. It seems reasonable to suppose that the reason for this is that humans have a tendency to try to act in accordance with what they believe to be rational. If we want to know why this is, or whether this human tendency is appropriate, we need to understand what it means for an action to be rational.

Thirdly, many philosophers have partially based their theories of morality on their notions of rational action or the closely related, if not identical, notion of what actions lead to what is good for or what is in the self-interest of an individual. Let me list five examples: 1) Kant argued that morality is a rational constraint on action and that it, therefore, always is irrational to act contrary to morality. 2) Utilitarians have argued, based on the notion that *what it is rational for an individual to do is to perform the action which best helps maximize her utility*, that what it is rational for a society to do is to carry out the policies which maximize total utility among the people of the society. 3) John Rawls, and other contractualists have argued that the laws of morality are what would be decided upon by self-interested individuals in the hypothetical scenario which Rawls described as the original position under the veil of ignorance, where they all have to voluntarily agree to rules for a society in which they will have to live. 4) David Gauthier, a prominent contractarian, based his notion of morality on an analysis of prisoner's dilemma cases. Prisoner's dilemma cases are cases where everyone acting out of a principle of pure self-interest would lead to a worse outcome for every single individual, compared to everyone following certain rules, where these would require them to abstain from performing certain actions out of pure self-interest.

Gauthier identified morality as the rules which self-interested individuals bargaining with each other in prisoner's dilemma cases, out of self-interest would agree to. 5) The last example is from the teachings of Aristotle. He thought that what is good for an individual is to perform actions that help the individual develop eudaimonia, and he also took these to be the actions it is moral for an individual to perform.

In order to assess whether it is reasonable to base an argument for a theory of morality on a particular notion of what rational action is, it is obviously crucial to consider how we should understand the concept of rational action. Furthermore, if rational action is not identical to '"action which is in the self-interest of an individual"', it is an open question which of them it is more reasonable to base a particular argument for a theory of morality on. It is important to know how to understand the concept of rational action, in order to be able to judge if arguments that are based on the latter concept should be based on the former instead or vice versa. To properly asses the reasonableness of an argument for a moral theory based on a particular notion of the concept of rational action or the concept of '"action that is in the self-interest of the actor"', it is, therefore, crucial to understand what rational action is.

My approach for answering the question of what rational action is is primarily based on an assessment of what interpretations best seem to correspond to our normal use of the terms 'rational', 'rationally required', and 'irrational'. In order to properly answer the question using this type of approach, it is necessary that one considers many of the interpretations philosophers have proposed with regard to how rational action should be understood. I will, through this thesis, discuss many such interpretations to understand and methodologically consider them in a process to identify an answer which seems reasonable.

Let me give a short overview of how this thesis is structured. Chapter 2 is centered around establishing why it seems reasonable to pursue an understanding of rational action which says that the only requirement an action needs to meet in order to be a rational action is the instrumental requirement. I do so by arguing for the plausibility of the instrumental requirement and the implausibility of The Kantian View. In Chapter 3, I explore how we should understand this requirement. I point out nine questions which can be asked with regard to how the requirement is to be interpreted, and multiple available options for answering each question. I then argue that we should choose the understanding of rational action which is the most in accordance with our use of terms related to practical rationality.

With the help of this method, I spend most of Chapter 3 considering which answers seem the most plausible. In Chapter 4, I propose a particular understanding of rational action based on the answers in Chapter 3 and consider whether it is compatible with certain ideas about rational action.

My criticism of The Kantian View is a negative argument, where I argue that his view seems implausible by presenting cases where the implications of his view seem very counterintuitive. When I compare different interpretations of The Instrumental Requirement, I also use example cases to make clear why some interpretations seem more linguistically plausible than others, and for some of the interpretations, I consider whether they entail that the concept of rational action has properties it seems reasonable to suppose that this concept must have. This discussion yields both the negative result that some interpretations seem less plausible, and the positive result that some interpretations seem more plausible. The main positive result of the thesis as a whole is the result that a particular view of what rational action is seems reasonable. I will come to call this view 'The End Point Pursuit Understanding of rational action '.

1.2 - Clarifications

Concepts that are not explored

There are certain concepts one may or may not think are connected to the concept of rational action, but which this discussion does not cover because, in this thesis, I do not assume that an understanding of these concepts is required to understand the concept of rational action. In order to make clear what some of these concepts are, let me contrast them with some concepts that will be the center of discussion:

Firstly, questions of practical rationality are going to be discussed, while questions regarding theoretical rationality will not be considered. Philosophers distinguish between what is called 'practical rationality' and what is called 'theoretical rationality'. Instead of going into a detailed discussion about the differences between the two, it will suffice to say that theoretical rationality is concerned with the beliefs of a person, and practical rationality is concerned with the actions or intentions of a person. The topic of discussion of this thesis

is on the topic of practical rationality only, and I will not explore the question of what theoretical rationality is.

Secondly, only the question of what a rational action is will be discussed, not the question of what a rational person is. In this thesis, I am exploring what a rational action is, without assuming that we need to base our understanding of what a rational action is on an understanding of what a rational person is. For instance, I do not assume during the discussion that the concept of '"a rational action"' is identical to '"an action performed by a person while being a rational person"'. It might namely be that the concept rational action and the concept '"rational person"' are entirely independent of one another, meaning that neither the concept of '"rational person"' is needed to make sense of rational action nor vice versa. I will explore what a rational action is only by relying on means which are not making references to the concept of '"rational person"'. In fact, in this thesis, I do not mean to say anything about what a rational person is.

Now, the term 'requirements of practical rationality' conceivably can be used to include requirements that a person needs to fulfill in order to be a rational person, such as a requirement *to have a coherent set of intentions*. However, when I use the term from here on out, I do not mean to refer to these types of requirements. Instead, I will use the term only to refer to the requirements which an action needs to fulfill in order to be a rational action.

Thirdly, in my discussion, I do not make references to the concept of '"having a reason to perform a given action"'. '"Having a reason to perform a given action"' is commonly thought to be closely tied to the concept of rational action. While I think it is likely that one of these concepts can be defined based on the other, I do not think it is necessary to mention the term '"having a reason to perform a given action"' in order to discuss what makes an action rational. To me, '"having a reason to perform a given action"' seems like a concept which is harder to get a grip on than the concept of rational action, and because of that, it might make a discussion of rational action less clear. Since I do not mean to say anything about the concept of '"having a reason to perform a given action"' or the relation between this concept and rational action, I find it best to avoid talk of the former altogether.

Punctuation scheme

I will use the following punctuation scheme:

- A word, term, or sentence fragment that stands between two single apostrophes refers to the word, term, or sentence fragment itself. Example: ‘Term’.
- A term or sentence fragment that stands between two triple apostrophes refers to the concept the term or sentence fragment refers to. Example: ‘“Term”’.
- What stands between two double apostrophes are quotations. Example: “This is a sentence fragment”.
- What stands between two single asterisks are statements, claims, propositions, or descriptions of beliefs, requirements, desires, actions, and the like. Example: *This is a claim*. It is meant to make sentences easier to read by signaling what is part of the statement and what is not.
- What stands between two double asterisks, are terms which are unusual in the sense that I use them to refer to something it is not commonly used to refer to, or because the term is new. Example: **Term**. What they refer to will be clear from the context where they are used.

‘Term’ and ‘word’

For the purposes of this thesis, what I call ‘term’ and what I call ‘word’ are slightly different. A given word is, in this thesis, a particular combination of letters, where the combination of letters can be used in one way to refer to one thing, and in another way to refer to another thing. An example of this is the letter combination ‘watch’ which can be used to refer to *a device that shows time* or can be used to refer to *the act of looking at something in an attentive way*. A given term, on the other hand, is in this thesis a particular use of a particular word. For instance, when ‘watch’ is used to refer to *a device that shows time*, this is a different term from when ‘watch’ is used to refer to *the act of looking at something in an attentive way*.

The difference between ‘rational’ and ‘rationally required’

That a given action is rational and that a given action is rationally required mean different things, in my discussion. If it is rational for an individual to perform action X, it means that

the individual's action is rational *if* it is action X. If a given action is *rational* for an individual to perform in a given set of circumstances, it does not exclude the possibility that there could be other actions that it would be rational for the individual to perform in those circumstances. If it is rationally required of an individual to perform action X, it means that the individual's action is rational *only if* it is action X. This entails that when a given action is rationally required for an individual to perform, and the individual is not performing this action, then she would necessarily be performing an irrational action.

2 - Why it seems reasonable to pursue a Humean view

2.0 - Chapter introduction

One view which seems plausible is the view that in order for an action to be rational, it needs to meet an instrumental requirement. This is called The Humean View of Practical Rationality. The most prominent alternative to this view is The Kantian View of Practical Rationality, which is the view that in order to be a rational action, an action needs to meet such an instrumental requirement, but also a categorical requirement.

This chapter is divided into two sections. In Section 2.1, I argue why it seems reasonable to think that what I call The Instrumental Requirement is a requirement an action needs to fulfill in order to be rational, and I offer a formulation of this requirement. In Section 2.2, I provide an account of what I will call The Categorical Requirement, and I argue that it has very counterintuitive implications and, therefore, seems implausible. I conclude that it because of this seems reasonable to pursue a theory of practical rationality which is based only on The Instrumental Requirement.

2.1 - The Instrumental Requirement of Practical Rationality

Kant claims that “whoever wills the end also wills (insofar as reason has decisive influence on [his] actions) the indispensable necessary means to it that are within his power" (Kant, G 4, p.414). It is commonly thought among philosophers, that in order for an action to be rational, it must at least meet a particular requirement of the type above, though there are disagreements regarding the exact details of this requirement. I will refer to this requirement as ‘The Instrumental Requirement of Practical Rationality’ or ‘The Instrumental Requirement’ for short. In the next paragraphs, I will argue for a particular formulation of The Instrumental Requirement. It is important to note that the understanding of The Instrumental Requirement which I propose, is significantly different from that of Kant just mentioned. Furthermore, the formulation I arrive at here is approximate in the sense that it is

open to different interpretations. Examining different possible interpretations of the requirement will be the objective of the following chapters.

To start off this analysis of what it means for an action to fulfill The Instrumental Requirement, consider the following case:

Example: The Chess Case

Imagine that an individual who is in the middle of a chess game, desires to achieve goal W of winning the game, that she correctly believes that achieving goal W also would lead to something which is good for her, that she correctly believes that her achieving goal W would lead to something which is independently good, and that she has decided that she is going to pursue goal W. For the sake of illustration, let us also suppose that achieving this is all the individual desires, that it is the only thing that would lead to what is good for the person, that it is the only thing that leads to what is good per se, and that it is the only thing she intends to pursue. Imagine that, at this point in the game, the individual only has two valid moves, **move A** and **move B** to choose between, and that she is contemplating which of the moves to perform. Suppose that she correctly believes that move A is terrible and most definitely will lead to a loss, while move B is excellent and that she will most likely win the game if she performs it. Suppose that she, despite this, for no good reason at all, performs move A and because of it loses the game.

Performing move A under these circumstances seems to be a clear case of irrational action, while it seems that performing move B in the case above would have been rational. It is commonly thought among philosophers, that the reason why some actions are rational while others are not is that there are certain requirements, that are such that an action is rational if and only if it meets all of them, and consequently that an action is irrational if it fails to meet any of them. The Instrumental Requirement would be an example of a requirement one can propose as being among these requirements. There is disagreement about what exactly these requirements are, and whether there are several requirements or whether there really is just one. Regardless, the general thought that whether an action is rational or not entirely depends on whether it meets specific requirements seems like a very sensible and promising approach for analyzing what makes it so that some actions are rational while others are not. I,, therefore, ,presuppose that there indeed are such requirements and that if we are able to pin down what these requirements are, we will have determined what makes an action rational. Given this, the appropriate task of any theory of what rational action is, is to try to determine

exactly what these requirements are. I am going to refer to such requirements as 'requirements of practical rationality'.

Constructiveness

Given that there are specific requirements which an action must meet in order to be rational, and given that move B in the example above is a rational action and move A is not, it must be the case that move B meets all of these requirements, while move A fails to meet at least one. This entails that there must be at least one difference between move B and move A, which causes move B to meet such a requirement and causes move A to fail to meet it. It seems reasonable to think that such a relevant difference is the fact that move A does not help the individual achieve goal W while move B does. We can say that move B is constructive for achieving goal W while move A is not. It, therefore, seems that, in the case above, move A's and move B's constructiveness with regard to achieving goal W determines whether or not the action is rational, where constructiveness is the degree to which each move helps the individual achieve goal W. From this observation it seems reasonable to think that a goal can be what we can call 'a rationally relevant goal', where an action's constructiveness with regard to achieving a rationally relevant goal plays a part in determining whether or not the action is rational.

Before I continue the discussion, some clarifications are in order. What exactly makes a goal a rationally relevant goal is not a given, and a number of different proposals have been made with regard to what makes a goal a rationally relevant goal.

It is common to think that a goal is rationally relevant only if the actor intends to achieve it. Some philosophers think that in addition to this, a goal is only a rationally relevant goal if achieving it is good per se (Kolodny & Brunero, 2018, p.5). Others think that it is relevant only if one desires to achieve it (Kolodny & Brunero, 2018, p.4). One might alternatively think that it is only relevant if it is good for the actor if the goal is achieved. A fourth possible view is to think that the only thing required for something to be a rationally relevant goal is that the actor intends to achieve it. At this point in my discussion, I will presuppose that one of these views is correct, but not make any assumptions as to which one. In the case above, goal W fulfills all of these requirements, and regardless of which of the four views we think is correct, goal W will be a relevant goal for determining whether or not an action is rational.

Secondly, while there only is a single goal token in the case above, namely *to win*, it is possible that there can be more than one goal token in other cases. While recognizing this possibility, I will for simplicity write ‘the rationally relevant goal’ also in cases where one might think that there could be multiple tokens of rationally relevant goals, and 'the rationally relevant goal’ can unproblematically be read as ‘the rationally relevant goal, or goals if we think that there can be more than one rationally relevant goal’.

Thirdly, it is a question whether it is *the actual constructiveness of actions* which determines whether or not they are rational to perform, whether it is *how constructive they are believed to be by the actor* which determines this, or whether it is *how constructive the evidence evidence held by the actor suggests they are* which determines this. I will leave it open for now which view we should take. Although, again for the sake of simplicity, I will write ‘the constructiveness of an action’, when I mean *the constructiveness of an action, how constructive the actor believes the action to be, or how constructive the evidence of the actor suggest the action is*. Having made these things clear, I will now go on with the discussion.

Given the fact that it seems that the constructiveness of an action for achieving the rationally relevant goal determines which actions are rational, it may seem tempting to think that one requirement of practical rationality is that an action is constructive for achieving the rationally relevant goal, where an action fails to meet this requirement if it is not constructive and it meets the requirement if it is constructive. However, as the next paragraphs will show, that this is a requirement of practical rationality does not seem to be entirely accurate.

The meaning of 'rational action’ is based on relative constructiveness

We would not typically say that an action is rational if there is a more constructive action for achieving the rationally relevant goal. To see this, consider the cases below:

Example: The Passerby Case - Version 1

Imagine a case where an individual is out on the streets asking for money and has the rationally relevant goal of collecting as much money as possible. Imagine that the passersby systematically are giving a lot of money to the person when she is sitting still on the ground begging and are systematically giving a smaller amount of money when she is playing the violin.

In such a case, it does not seem intuitive to say, given that collecting as much money as possible is the rationally relevant goal, that it is rational for the individual to play the violin if she only has the two options available. This, despite the fact that playing the violin, also is constructive for getting passersby to donate money. Hence, it does not seem that an action being constructive for achieving the rationally relevant goal in itself is enough for us to say that it is rational for an individual to perform the action, given that all other requirements of practical rationality are fulfilled.

What seems to be enough, given that all other requirements for being a rational action are fulfilled, is that an action is more constructive for achieving the goal than all the other actions available. This seems clear when we compare the case above with the following case:

Example: The Passerby Case - Version 2

Imagine a scenario like the one in the example above, where playing the violin yields the same amount of money, but where the individual does not have the option of sitting still on the ground. Imagine that she instead has the option to play the guitar, but that playing the guitar would systematically make the passersby donate less money than playing the violin would.

In this scenario, it seems that, given that collecting as much money as possible is the relevant goal, we would say that it would be rational to play the violin if the two options were all the options available. The only relevant thing that seems to have changed between the two cases is that in Version 2, where we would say it is rational to play the violin, the only other option available is a comparatively less constructive option for achieving the rationally relevant goal, while in Version 1, where we say that it is not rational to play the violin, the only other option is a comparatively more constructive option for achieving the rationally relevant goal. This seems to suggest that it is an action's constructiveness *compared* to the other available options, and not its constructiveness per se, determines whether it is rational to perform the action or not.

It, therefore, seems reasonable to think that when we say that it is rational to perform an action, something we mean to say that the action is at least is as constructive for achieving the rationally relevant goal, compared to all the other actions available, (where the act of doing nothing also counts as an action). For instance, if we say that *if one has the goal of doing well in school, then it is rational to study* it seems reasonable to think that we

commonly at least mean to say that *studying is an action which is at least as constructive as any other action for achieving the goal of doing well in school*. It, therefore, seems that it is a requirement for *an action to be rational for an actor to perform* that *to perform the action is at least as constructive, for achieving the rationally relevant goal, compared to all the other actions available to the actor* or as I will often say *is the action which best helps achieve the rationally relevant goal of the actor*.

I will, therefore, take The Instrumental Requirement to be the following:

The Instrumental Requirement of Practical Rationality

In order for an action X to be rational for an individual to perform, given that goal A is the rationally relevant goal, it is a requirement that action X is at least as constructive for achieving goal A compared to all other available actions.

The above formulation entails that when all other requirements of practical rationality are fulfilled, it is rational to perform an action X if it is at least as constructive as all other available actions for achieving the rationally relevant goal. As will become evident in the next chapters, the exact meaning of the above formulation of The Instrumental Requirement is open to many different interpretations.

2.2 - The Categorical Requirement of Practical Rationality

Kant's Categorical Imperative

While Kant fully recognizes that The Instrumental Requirement plays an important part in determining which actions are and which actions are not rational, he argues that there also is a second type of requirement of practical rationality, which also plays a crucial part in determining this. This requirement is reflected by what Kant calls 'The Categorical Imperative'. The following statement, commonly referred to as 'The Universal Law Formulation' is one of the formulations of The Categorical Imperative which Kant provides:

The Universal Law Formulation

Act only in accordance with that maxim through which you can at the same time will that it become a universal law. (O'Neill, 2004, p.98)

Three of the terms of this formulation need clarification. Firstly, by the term ‘maxim’, Kant means a general principle for acting which an individual can adopt, where an individual adopting a maxim means that she has decided that she intends to let her actions follow that maxim (O’Neill, 2004, p.94). For instance, one can, in principle, adopt the maxim of *eat cake only on Sundays*, and on the basis of having adopted this maxim, decide not to eat cake at a café during a weekday. It is up to each individual to decide which maxims to adopt (O’Neill, 2004, p.94). Christine M. Korsgaard writes that to adopt a maxim is in the Kantian picture to commit “[oneself] to realizing the end” the maxim is referring to, and that it means the same as to will this end (Korsgaard, 1997, p.57). In other words, committing oneself to realize an end is the same as adopting the corresponding maxim. This type of commitment can be thought of as something close to having an intention to realize the end (Kolodny & Brunero, 2018, pp.45-49). This entails that if one has adopted the maxim *eat cake only on Sundays* then one has committed oneself to act in a way that realizes this end. The Universal Law formulation hence reflects a restriction on what maxims it can be rational to act based on, or in other words, what it can be rational to commit oneself to bring about. According to Kant, it is only rational for an individual to perform an action if that action is based on a maxim that does not violate The Categorical Imperative. Secondly, *a maxim which is a universal law* Kant defines as a maxim which all actors have permanently adopted. For example, if the maxim *eat cake only on Sundays* actually were a universal law, it would mean that all individuals had committed themselves to realize the end of the maxim. Something which is important to emphasize is that The Universal Law formulation is not a claim that there exist maxims that have become universal laws. What Kant *is* claiming is that it is a requirement of practical rationality that one *can will,* in a hypothetical sense, that the maxims which an action is based upon is adopted by everyone (O’Neill, 2004, p.98). This requirement is what The Universal Law Formulation reflects. Thirdly, according to Kant, one can will something if an only if one can coherently will it. ‘Will’ in The Universal Law formulation can, therefore, be read as ‘coherently will’.

Kant provides a handful of other formulations of The Categorical Imperative intended to reflect the same type of requirement of practical rationality as each other, and many of the formulations all sound very different from one another. Despite this, he claims that they are

all equivalent to each other and that they are really just different formulations of the same underlying requirement of practical rationality (O'Neill, 2004, p.98). When she explains how two of the formulations of The Categorical Imperative can in some sense be equivalent, O'Neil writes: "Both formulas state procedures for rejecting maxims whose universal adoption would undermine at least some [other individuals'] possibilities of like action" (O'Neill, 2004, p.106). These are the maxims one cannot coherently will to be universally adopted according to Kant. Each of the formulations acts as a way to determine which types of maxims are such that one cannot coherently will their universal adoption. The essence of The Categorical Imperative is that an action is not rational if one cannot coherently will the universal adoption of the maxim it is based on. As we have seen, this is precisely what The Universal Law expresses.

The Universal Law Formulation *act only in accordance with that maxim through which you can at the same time will that it become a universal law* obviously means the same as *do not act in accordance with the maxim through which you cannot at the same time will that it become a universal law*. Given all of what is said in these paragraphs, it follows that The Universal Law Formulation is identical to the following requirement, and this requirement seems to be the central requirement which all of the formulations of The Categorical Imperative reflect:

The Categorical Requirement of Practical Rationality

It is a requirement of practical rationality to not act on the basis of maxims which one cannot coherently will to be adopted by everyone.

It, therefore, seems that we should understand Kant as proposing that The Categorical Requirement of Practical Rationality, or what I will call 'The Categorical Requirement' for short, is a second requirement of practical rationality in addition to The Instrumental Requirement.

Now, Kant thinks that there are some maxims which one cannot will to be adopted by everyone simply because it would be impossible, even in principle, for everyone to adopt such a maxim, and this being the case, willing such a maxim to be adopted by everyone would, therefore, not be a coherent thought. Such maxims "cannot, therefore, be thought of as principles" which everyone can adopt (O'Neill, 2004, p.103).

Kant thinks that a maxim of false promising is an example of a maxim which would violate The Categorical Requirement for this reason. As O'Neil writes: "[Universal false promising would include] the destruction of trust" (O'Neill, 2004, p.99). Without people trusting promises, making false promises successfully becomes impossible. Because the universalization of false promising would entail that false promising becomes impossible, universalization of false promising is an impossibility. Similarly, because lying and any other kind of deception of others destroy trust, its universalization would make it impossible to lie or deceive successfully. To act on the basis of maxims such as a maxim of false promising, a maxim of lying, or a maxim of deception violates The Categorical Requirement for the following reason: "[Such principles] cannot coherently be willed as universal laws because their universal adoption (*per impossible*) would predictably undercut the possibility of adopting those very principles for at least some others" (O'Neill, 2004, p.102).

Why The Categorical Requirement has implausible implications

The Categorical Requirement seems to have very implausible implications. I am now going to present three cases that I take to clearly illustrate this. Consider first the following case:

Example: The Kidnapped Case

Suppose that an individual has been kidnapped, and her kidnapper tells her that she can go to the store by herself if only she promises not to escape. Suppose that the kidnapper is sincere and that the kidnapped individual believes that the kidnapper would let her go to the store alone if she promises not to escape and that she certainly would be able to escape if she went to the store alone.

Making the false promise would undermine the possibility for people to make false promises in the future, and this action is, therefore, not an action that can be universalized. Because of this, The Categorical Requirement would imply that it is irrational for the individual to make the false promise in The Kidnapped Case. This seems intuitively implausible. In fact, it seems natural to think is that it would clearly be rationally required of the individual to make the false promise in the case above. Now, consider the second case:

Example: The Trojan Horse Case

Imagine a scene from the Greek legend of the Trojan Horse. The Greeks and the city of Troy are at war. The Greeks have besieged the City of Troy, but the city is well defended. The Greeks are running

out of resources and realize they cannot defeat Troy with brute force. However, they come up with a creative, strategical plan based on deception, and they are convinced that it will work. The first part of the plan is to trick the Trojans to bring greek soldiers into Troy by pretending that they are giving up and sailing away and leave a large wooden horse statue outside the gates of Troy, hoping that the Trojans will take it inside their walls not knowing that a group of soldiers is hiding inside the horse. The next part of the plan is for the soldiers inside the horse to come out at night and open the gates for the Greek army, which is hiding outside. According to the legend, the plan works, and by executing the plan, the Greeks successfully conquer Troy.

It seems intuitive to think that in such a scenario, it is rational for the Greeks to perform this strategy. However, The Categorical Requirement would imply that this, in fact, would be irrational. To defeat an enemy in this way is not something that can be universalized because if everyone were performing this strategy to defeat their enemies, then people would stop falling for the deception, and conquering enemies in this way would be impossible. Let us consider one more case:

Example: The Atomic Bomb Case

Say that a democratically elected leader of Country A, a nuclear power, has gone mad and wants to destroy his own country. Imagine that he wants to launch the entire nuclear arsenal of the country against another nuclear power, knowing that they will launch their full arsenal at Country A in return, which would destroy Country A completely, and kill all of its inhabitants. Suppose that he asks the woman next in charge to bring him the communication device he needs to use in order to give the order. Imagine that she knows that the leader has gone mad, and that she wants to stop the leader from causing their country to be destroyed, but that she knows that if she refuses to bring the leader the communication device, the leader will get it himself. Imagine that she tells the leader that she will get him the communication device, but that she instead returns with what she knows is a convincing but fake copy of this device, which she has had created in the past. Suppose that since the leader thinks it is the real device, he will spend time trying to give the order using it and that this gives the woman next in charge time to contact certain people and successfully organize with them in order to stop the leader from causing the destruction of Country A. Imagine that this is the only course of action available to the woman next in charge to stop the destruction of Country A.

Intuitively it seems that if anything is a rational action, it is rational for the woman next in charge to stop the leader by using these means. However, again The Categorical Requirement would imply that stopping the destruction of country A and the deaths of all of its habitants in this way would be an irrational action because the action is not universalizable. Similar to

what was the case in the other cases, if everyone acted in this way, then people in these situations would stop trusting their person next in charge to bring them such a device, and it would not be possible to stop a person in this way.

That The Categorical Requirement has implications like these, which seem very counterintuitive, makes it seem very unlikely that in order for an action to be rational, it has to meet The Categorical Requirement. It, therefore, seems reasonable to presume that The Categorical Requirement is not a requirement of practical rationality. Given that The Kantian View seems prima facie implausible it seems reasonable to pursue a theory of rational action based on The Humean View that an action is rational if and only if it meets The Instrumental Requirement.

3 - How to interpret The Instrumental Requirement

3.0 - Chapter introduction

In this chapter, I will start by pointing out nine questions which it seems reasonable to ask with regard to how The Instrumental Requirement is to be interpreted. I will, in Section 3.2 ,establish the reasonableness of to types of arguments I will rely upon later in the chapter. In Sections 3.3 through 3.11, I will discuss which of different available answers seem to most reasonable in response to each of the nine questions.

3.1 - A whole host of available interpretations

In Chapter 2, I argued that The Instrumental Requirement can be formulated as the following:

The Instrumental Requirement of Practical Rationality

In order for an action X to be rational for an individual to perform, given that goal A is the rationally relevant goal, it is a requirement that action X is at least as constructive for achieving goal A compared to all other available actions.

It turns out that one can take many different views about exactly how The Instrumental Requirement is to be understood. The different understandings will imply different things when it comes to the question of what actions it is rational to perform. It is not immediately obvious which of the understandings we should go with. I will now list nine different questions one can ask with regard to how the requirement is to be interpreted. For each question, there are multiple available answers. ‘Interpretation’ will, in this thesis, refer to a particular answer to one of these questions.

1. Whether the rationally relevant goal is determined by desires, prudence, independent value, or by what the actor actually intends

One can ask what a rationally relevant goal is exactly. Now, there seems to be general agreement that the rationally relevant goal is what it is rational for the actor to intend to achieve, and that if it is rational for an individual to intend to achieve something, then her action is rational only if she intends to perform what she believes to be the best or necessary action to achieve it, and she performs the action for this reason. John Broome gives an account of why it would be irrational not to intend to perform such an action. He writes:

> [S]uppose you intend to open the wine, and you believe that to do so you must fetch the corkscrew from the kitchen. This information and belief require you to intend to fetch the corkscrew from the kitchen. (Broome, 1999, p.412)

This entails that an action is always irrational if one intends to achieve something and does not intend to perform the action one believes to be the best or necessary mean to achieving it. Given that rational action at least partially is about intending the means to one's intended ends, this seems to suggest that a goal is not a rationally relevant goal unless the actor intends to achieve it. While this seems correct, this is not a complete answer for what a rationally relevant goal is because one might ask what determines whether a goal is rational or irrational to intend to achieve. In other words, what, if anything, determines the rationally relevant goal of an actor apart from the fact that the actor intends to achieve it.

At least four alternatives to answer this question have been pursued or might seem reasonable to pursue. One view is that the rationally relevant goal is determined by what is of independent value (Kolodny & Brunero, 2018, p.5). This view can be referred to as 'The Value-Based Interpretation'. A second view, which we can call 'The Desire-Based Interpretation', is that the rationally relevant goal is determined by what the individual desires (Kolodny & Brunero, 2018, p.5). A third view, which we can call The Prudence-Based Interpretation is that the rationally relevant goal is determined by "what is best for the individual on the whole" (Korsgaard, 1997, p.29). A fourth view, which we can refer to as 'The Actual Intention Interpretation, which might seem plausible, is to deny that anything apart from the fact that an actor intends to achieve something makes it a rationally relevant

goal and that the rationally relevant goal of an actor is whatever she actually intends to try to achieve in the moment she performs the given action.

I will come to argue that we have reason to take The Desire-Based Interpretation. If we believe that the desires of an actor determine her rationally relevant goal, one might ask whether all desires do or whether just particular desires do. The desires that are taken to determine this I will call ‘rationally relevant desires’.

Now, if we take The Desire-Based Interpretation, then all the questions that follow below can be asked.

2. Whether what one is rationally required to do is to do what best helps fulfill the totality of one’s rationally relevant desires

It might be that it often or usually is the case that one action is most constructive for achieving one particular rationally relevant desire and that another action is most constructive for achieving another. One might ask whether one for such cases should understand The Instrumental Requirement as a requirement to perform the action which best helps fulfill the totality of the actor’s rationally relevant desires, or if all The Instrumental Requirement only provides basis for saying is that with regard to fulfilling one of the desires one of the actions is rational, and with regard to achieving another desire another action is rational. The former view I will refer to as ‘The Holistic Interpretation and the latter view as ‘The Limited Role Interpretation’.

3. Whether only intrinsic desires are rationally relevant desires, or whether instrumental desires are too

We can distinguish between what we can call 'instrumental desires’ and what we can call ‘intrinsic desires’. Instrumental desires are desires which one has solely because achieving them helps achieve something else that one desires. For instance, one might have the desire of renting a motorbike, but only because one has the desire to go on a motorbike trip and that without one having the desire to go on a motorbike trip one would not have the desire to rent a motorbike. In that case, the desire to rent a motorbike is an instrumental desire. Intrinsic desires are desires which one does not have solely because their achievements are instrumental to achieving the fulfillment of some other desire which one has. For example, if it would still be a desire for an individual to go on a motorbike trip even if this did not

achieve anything else which she desires, this would be a genuine intrinsic desire. Based on this distinction, one might ask whether we should think that both intrinsic and instrumental desires can be rationally relevant desires or whether only intrinsic desires are. The view that both types of desires can be rationally relevant I will call 'The Instrumental Inclusive Interpretation', and the view that only intrinsic desires can be rationally relevant 'The Intrinsic Interpretation'.

4. Whether only intrinsic desires that cannot be criticized on the bases of conflicting with other intrinsic desires are rationally relevant desires

One can ask whether an intrinsic desire can be rationally irrelevant due to the fact that it can be criticized with reference to other intrinsic desires. The interpretation that this can make an intrinsic desire rationally irrelevant we can call 'The Relevance of Criticism Interpretation', while the view that it cannot we can call 'The Irrelevance of Criticism Interpretation'. To find an answer to the question, we need to consider different views for how an intrinsic desire can be criticized in such a way. One view is the view that an intrinsic desire can be criticized on the basis that it is conflicting with another intrinsic desire that is "subjectively more important to the agent" (Wallace, 2018, p.14). Another view is that an intrinsic desire can be criticized on the basis that having it overall hinders the actor from acting in accordance with practical rationality. If we find any of these views convincing, the question becomes whether we have grounds for thinking that intrinsic desires are not rationally relevant desires if they can be criticized in the way described by the given view.

5. Whether only the desires one has at the time of deliberation are rationally relevant desires or whether future desires are rationally relevant desires too

Given The Holistic Interpretation, it might seem natural to ask whether desires one will have at some point in the future can be rationally relevant desires. One can hold the view that only the desires which one has at the point in one's life when one is making the choice of what action to perform can be rationally relevant desires (Wallace, 2018, p.17). We can call this view 'The Contemporary Desires Interpretation'. An alternative view, which we can call 'The Future Inclusive Interpretations' is the view that in addition to these desires, all the desires which the actor will or is anticipated to have at some point in the future can be rationally relevant desires (Wallace, 2018, p.17).

6. What to take contemporary intrinsic desires to be

One can distinguish between what can be referred to as ‘occurrent desires’ and what can be referred to as ‘standing desires’. The following description from Schroeder gives a basic idea of the distinction: “Standing desires are desires one has that are not playing a role in one’s psyche at the moment”, and “[o]ccurrent desires […] are desires that are playing some role in one’s psyche at the moment" (Schroeder, 2017, p.22). Standing desires include desires which sometimes are occurrent, such as a desire for a pet, but which one has forgotten about in the moment or for the time being.

Based on this distinction, we can interpret The Desire-Based version of The Instrumental Requirement in at least two different ways which seem prima facie plausible. One view, which we can call ‘The Standing Inclusive Interpretation’, is that both occurrent and standing desires can be rationally relevant desires. Another view, which we can call ‘The Occurrent Desires Interpretation’ is that only occurrent desires can be rationally relevant desires. Suppose that in a given situation, the only occurrent desire one has is to drive safely to a given destination as fast as possible. According to ‘The Occurrent Desires Interpretation’, what would be rational to do in this situation would be to take the fastest route one knows. The Standing Inclusive Interpretation, on the other hand, could require an actor with only such an occurrent desire to perform other actions, such as driving by a restaurant to pick up food, given that this would have been the most constructive action for helping to fulfill the totality of all of her contemporary desires.

7. Whether only laundered desires are rationally relevant desires

One can ask whether intrinsic desires are rationally relevant only if they have passed some type of requirement we can call a ‘laundering requirement’, that is, a requirement that an intrinsic desire must have been produced in a particular way. A laundering requirement is meant to exclude desires which it does not seem that one has adequate reason, in some sense, for trying to fulfill. For example, a laundering requirement might be aimed at excluding desires which have been formed on the basis of false beliefs or at excluding desires which have been formed due to an individual being in a certain temporary state of mind, such as anger (Wallace, 2018, p.18). The view that there is a requirement of the first kind we can call

‘The Deception Laundering Interpretation’ and the view that there is a requirement of the second kind ‘The Mental State Laundering Interpretation’. These views do not exclude one another as there might be that both types of requirements are requirements for an intrinsic desire to be a rationally relevant desire. The view that there is no requirement with regard to how an intrinsic desire is produced we can call ‘The No Laundering Interpretation’. According to this view, any intrinsic desire which fulfills all other requirements is a rationally relevant desire no matter what emotional state produced the intrinsic desire or which beliefs it was formed on the basis of.

8. Whether The Instrumental Requirement requires an actor to do what she believes will best help fulfill her rationally relevant desires, what actually would best help fulfill these desires, or what her evidence suggests will best help fulfill these desires

One can distinguish between the view that *The Instrumental Requirement requires that the given actor performs the action which *actually* will best help fulfill the rationally relevant desires of the actor*, the view that *what is required by the requirement is that the given actor performs the action which *she believes* will best help fulfill her rationally relevant desires, and the view that *what it requires the given actor to do is to perform the action which *the evidence* held by the relevant individual suggest is the action which will best help fulfill the rationally relevant desires of the actor* (Kolodny & Brunero, 2018, p.4). The first view I am going to refer to as ‘The Objective Interpretation’ and the second as ‘The Subjective Interpretation’. One can hold different versions of the third type of view based on whom one takes the relevant individual to be. One view is that this is the actor. Another is that this is the person making the claim about what is rationally required of the actor. And a third view is that this is the person who is considering such claims (Kolodny & Brunero, 2018, p.4). The first of these views I will call ‘The Evidence-Centered Interpretation’. This view seems like the most natural alternative to The Objective Interpretation and The Subjective Interpretation, and later, I will only spend time comparing the favorability of these three views.

9. Whether The Instrumental Requirement can be met through changing one's goal

One can ask if we should understand The Instrumental Requirement as a requirement to perform actions which serve one's rationally relevant goal, *or* if we should understand it as a more general requirement to make it so that one's rationally relevant goal and actions are coherent with one another (Wallace, 2018, p.12). The latter view we can call 'The Wide-Scope Interpretation' and the former 'The Narrow-Scope Interpretation'. According to The Wide-Scope Interpretation, The Instrumental Requirement can be fulfilled by an individual not only by her performing the actions which are most constructive for realizing her rationally relevant goal but also by her instead changing her rationally relevant goal. The difference can be illustrated by looking at the following case:

Example: The Swimmer Case

Let us suppose that there is an individual who has the rationally relevant goal of becoming a world-class swimmer, correctly believes that practicing is necessary to reach the goal.

In this case, The Narrow-Scope Interpretation suggests that that the person would be rational only if she practices, while The Wide-Scope Interpretation instead suggests that the person would be rational if she either practices *or* if she makes it so that becoming a world-class swimmer is no longer her rationally relevant goal.

3.2 - Approach for choosing between interpretations

All of the interpretations above seem prima facie plausible. However, we need to settle on one understanding of The Instrumental Requirement. In order to find out which interpretations to accept, we need a method that makes us able to pick interpretations over others in a principled way.

Choosing a concept of rational action means choosing how we understand the term 'rational action'. It seems clear that the degree to which an understanding of a term corresponds to how the term is normally used when it is used in a literal meaning by competent speakers of the given language can vary. For instance, the understanding of the term 'living' as referring to and only to '"the property of being something which sometimes moves"', seems to

correspond poorly to how we commonly use the term as there are so many things that move that we would not commonly characterize as living. When we use the term 'living' to describe something, we commonly do not merely want to say that it sometimes moves. It is clear that we normally want to say something different.

It seems reasonable to think that for most terms, there will be at least one coherent concept that corresponds perfectly to how the term is used. The idea is that such concepts will capture the essence of what language users normally want to say when using the term in a direct way. Take the term 'knowledge'. Philosophers have argued that knowledge is *a belief, which is true, which is justified, and which one has come to believe in only based on other justified beliefs*. If language users normally call something 'knowledge' if and only if these four conditions apply to it, then this concept corresponds perfectly to how the term is used.

I will use the term 'linguistic plausibility' to talk about how well or poorly a proposed understanding of a term corresponds to what language users typically want to express when using the term. If one of the interpretations of The Instrumental Requirement has implications that contradict what it seems clear that we commonly would characterize as a rational action or as an irrational action or has implications that imply that rational action lacks features which it seems that it must have I will call the understanding linguistically implausible. If a proposed concept of rational action clearly does not seem to be what we commonly mean to talk about when we use the terms 'rational' or irrational to describe actions it seems clear that it should be rejected as an understanding of rational action on the basis that it does not make sense of what we mean to talk about using the term. It is not that such a concept necessarily will be a useless concept in itself. It would be possible that the proposed concept is a useful concept in its own right. It just does not correspond to the term rational action in the way it is commonly used.

Now, what interpretations we choose for The Instrumental Requirement directly influences how linguistically plausible the resulting concept of rational action is. If a given interpretation results in a concept which seems linguistically implausible, I will say that the interpretation seems linguistically implausible. One can test how linguistically plausible an interpretation of The Instrumental Requirement is in multiple ways. One way which I will often be utilizing in the discussions in this chapter is to consider example cases where it may seem clear what we commonly would say it would be rational or irrational for an individual

to do, but where one of the proposed interpretations has implications for what is rational or irrational which are not corresponding with this. Another way is to identify some property that it seems reasonable to think that a concept of rational action must have and see whether a given interpretation entails that the resulting concept lacks this property. One can say that such a property is an essential property of the concept of rational action and that it is a requirement of an understanding of rational action that it entails that rational action has this property in order to be linguistically plausible. Both of these approaches will play a role in my discussion of which interpretations of The Instrumental Requirement we should adopt, which now follows.

3.3 - Desires, prudence, independent value, or what the actor actually intends?

As noted, The Value-Based Interpretation, The Desire Based Interpretation, The Prudence-Based Interpretation, and The Actual Intention Interpretation are the views that the relevant goal of The Instrumental Requirement is based in part on what is of independent value, the relevant desires of the actor, what is best for the actor on the whole, and what goal the actor actually intends to achieve by performing a given action, respectively. I will now go through them one by one to consider which of them seems the most favorable.

An argument from linguistic plausibility against The Value-Based Interpretation

Kolodny and Brunero uses the following example to illustrate how The Value-Based interpretation and The Desire-Based Interpretation have different implications for what is rational to do: "Suppose [a] madman's [only] desire is to set off a nuclear war […] [and that] the madman [correctly believes] that to press [a particular] button [will set] off a nuclear war, [while not doing it will not]" (Kolodny & Brunero, 2018, p.5). Given that the desire is a rationally relevant desire, it would, according to The Desire-Based Interpretation, be irrational for the madman not to push the button. Given that *not setting off the war* leads to outcomes of more independent value than *setting of the war* leads to The Value-Based Interpretation will imply that it would be irrational for the madman to push the button.

In order for The Value-Based Interpretation to be linguistically plausible, it must be the case that when an individual in a correct manner calls an action 'rational' she must be thinking that the action best helps achieve something which is of independent value. However, it seems that there are cases where we would commonly say that an action is rational to perform, but where if it is true that that action is the one which best helps achieve something which is of independent value, it is not easy to see what this thing which has independent value plausibly could be. This makes the notion that we commonly say that an action is rational only if we think it best helps achieve something which is of independent value seem linguistically implausible. Consider the following example:

Example: The Poker Case

Imagine an individual playing a game of poker against her personal computer, and the individual's only desire and intention is to win that game. She correctly believes that if she wins, she will neither become happier nor less happy compared to if she loses. Suppose that she finds herself in a decisive moment of the game. She is at the end of a round that she knows she needs to win in order to win the game and that she knows she is sure to lose the game if she does not. Imagine that she has the option to **fold**, which means forfeiting her opportunity to win the all-important round, and thereby be sure to lose the game or to **check**, which means staying in the game without any cost. Suppose also that she is aware that she has the best hand in the game, namely a royal street flush, and correctly believes that she will win the entire game of poker if she **checks**.

It seems that we would commonly say that **checking** would be the rational action for the individual to perform in the case above. Given this, in order to maintain that The Value-Based Interpretation is linguistically plausible, one would have to claim that we commonly would say this is because we would commonly believe that **checking** in the case above would best help achieve something of independent value. However, it is not easy to think of something which commonly would be said to have independent value which **checking** would best help achieve. For instance, things that have been frequently suggested as being independently valuable historically include *moral good as such*, *happiness as such*, and *people following God's will as such*. It seems unlikely that we would commonly think that the act of **checking** in the case above would best help achieve any of these things. Unless one can show that there is something which humans commonly agree is of independent value,

which can explain our use of the term 'rational action' for cases such as the above, it seems that we should think that The Value-Based Interpretation is linguistically implausible.

An argument from linguistic plausibility against The Actual Intention Interpretation

Chrisoula Andreou suggests that it is not irrational for an individual to perform an action which is necessary to achieve a goal unless she intends to achieve it (Andreaou, 2006). In arguing against The Desire-Based Interpretation she writes:

> [S]uppose an agent has a desire (or inclination or motivation) to X but no intention to X. Suppose that to X the agent must Y. Suppose finally that the agent does not Y. Is this a genuine failure on the part of the agent? I think not. (Andreaou, 2006, p.321)

However, The Actual Intention Interpretation seems to have linguistically implausible implications. To see this, consider the following case:

Example: The Drawing Case

Imagine a scenario where a child shows her mother a drawing she made and asks whether the mother thinks the drawing is any good. Let us suppose that the mother does not think the drawing is any good. Suppose that the mother has only two options available: A) Express her honest judgment about the drawing, and B) tell the child that she thinks that the drawing is good. Now, suppose that the mother only has two desires: 1) A strong desire for her child to be as happy as possible, and 2) a weak desire to express her honest aesthetic judgment about the drawing to her child. Suppose also, that the mother correctly believes that the only difference between action A and action B in terms of effects on the child's happiness would be that action A would cause her child to be a lot sadder for an extended period of time than action B would. Imagine that the mother, because of this, correctly believes that the child would be less happy if she expressed to her child her honest judgment. Say that the mother, despite this, only intends to express her aesthetic judgment about the drawing to her child.

In this scenario, The Actual Intention Interpretation suggests that it would be rational for the mother to tell her child her honest opinion about the drawing. This seems linguistically implausible. Given that her desire for the child to be as happy as possible is much greater than her desire to tell her honest opinion about the drawing, and that she only has these two desires it seems that we would commonly say that it would be irrational for the mother to tell

her honest opinion about the drawing. This seems to be a counterexample to Andreou's claims, and it suggests that The Actual Intention Interpretation is linguistically implausible.

Comparing The Prudence-Based Interpretation and The Desire-Based Interpretation

The difference between The Prudence-Based Interpretation and The Desire-Based Interpretation is more subtle due to the fact that it seems that individuals normally have a strong desire to do what leads to what is best for them on the whole. However, an individual may not necessarily have such a desire, and it is not necessarily the case that this is the only desire an individual has, because of this the implication of The Prudence-Based Interpretation and The Desire-Based Interpretation may sometimes diverge. The following example helps illustrate this:

Example: The American Football Case

Imagine a case of an individual who is really good at American football, and who plays it professionally. Suppose also that what is best for the individual overall is to stop playing American football due to the risk of head injury. Furthermore, suppose that the individual has no desire whatsoever to avoid getting a head injury, and his only desire is to live the lifestyle which his playing of American football leads to and that he has no other paths available to live that same lifestyle.

In such a scenario, The Desire-Based Interpretation would suggest that the individual is rationally required to keep playing American football, while The Prudence-Based Interpretation would imply that it is rationally required of the individual to stop playing American football. Now, because it can be hard to imagine an individual in a real-world scenario who truly does not desire to do what he believes is best for him after careful contemplation, one might make the assumption that doing what is best for oneself overall is the only thing any individual truly wants. If this turned out to be true, there would be no difference in implication between The Prudence-Based Interpretation and The Desire-Based Interpretation in real-world scenarios. However, as Korsgaard points out, such an assumption might be unwarranted:

> [E]mpiricist philosophers and their social scientific followers have obscured the difference between [a requirement of practical rationality based on The Desire-Based Interpretation and

> one based on The Prudence-Based Interpretation] by making the handy but unwarranted assumption that a person's overall good is what he "really" wants. (Korsgaard, 1997, p.30)

Given that it is not necessarily the case that The Prudence-Based Interpretation and The Desire-Based Interpretation will for real-world cases always imply the same when it comes to what actions it is rational of an individual to perform, there is a need to determine which of these interpretations we should adopt.

An argument from linguistic plausibility in favor of The Prudence-Based Interpretation

Those favoring The Prudence-Based Interpretation, might argue that this view seems more in line with our language than The Desire-Based Interpretation. A reason for why this might seem to be the case is that we commonly say things like *it would be rational for you to save for retirement even if you do not happen to have any desires right now for any of the things saving for retirement would help achieve*. The Desire-Based Interpretation would, of course, imply that it cannot possibly both be true that 1) it is rational for an individual to perform an action, and that 2) the individual does not have any desires of any kind which performing the action would help fulfill. At first glance, this seems to suggest that the theory that an action is rational only if it is what best helps fulfill one's rationally relevant desires is out of line with how we use the term 'rational action'.

The Prudence-Based Interpretation, on the other hand, might at first glance seem to be more in line with how we speak. We might think that in the example above, the speaker typically would think that it will be what best serves the overall good of the actor to save for retirement. If this is the case, this claim will not be at odds with The Prudence-Based Interpretation. Given that it is true that saving for retirement would be what best serves the overall good of the person, it would namely be true that it would be irrational for the individual to not save for retirement, according to this view.

However, The Desire-Based Interpretation might not be as at odds with this usage of the term 'rational action' as it first might seem. There are, namely, two ways which The Desire-Based Interpretation could correspond to the fact that we commonly think that statement like the above can be true:

One possibility is that it can be true due to the fact that the actor will have some desire in the future that the action we claim is rational would help fulfill. For the case above, it might, for

example, be true that the actor will have some desire in the future to be well off after retirement. Given this, we might explain that the statement could be true due to the fact that it would help fulfill desires the actor would come to have in the future, and not because it would be what best serves the overall good of the actor.

A second possibility becomes apparent when we notice, as I pointed out in Section 3.1, that we can distinguish between occurrent desires and standing desires, which we can think of as "desires that are playing some role in one's psyche at the moment" and "desires one has that are not playing a role in one's psyche at the moment" respectively. [p.22 Schroeder] It, namely, might be that, usually, when we call an action rational in cases similar to the example above, we are actually presuming that the actor has some standing desire which performing what we claim is rational would help fulfill. The speaker in the example above might, for example, think that *the actor has a strong standing desire to do what would make things better for her when she retires*, and that *saving for retirement is the action which best helps fulfill this standing desire*.

This would mean that claims like *it can be the case that it would be rational for you to perform a given action even though you do not currently desire any of the things performing the action helps achieve* normally are referring to occurrent desires only. Given this, in the example above, the claim might, therefore, be read as *it would be rational for you to save for retirement even if you do not happen to have any *occurrent* desires for achieving the things this helps achieve*. If 1) the claim of the example only is meant as a claim about occurrent desires, 2) it is true that the actor has a rationally relevant standing desire, and 3) saving for retirement is the action which best helps fulfill this desire, then the claim that *it would be rational for the actor to save for retirement even if she does not happen to have any occurrent desires for any of the things performing the action helps achieve*, would be correct according to The Desire-Based Interpretation, given that the actor has no other rationally relevant desires.

That we would commonly insist that it would be rational for someone to perform a given action even when we do not think they have any occurrent or standing desire that the action would help fulfill seems far less obvious. For example, let us suppose that a person does not have any desire, neither occurrent nor standing desire that saving for retirement would help fulfill. A reason for this might, for instance, be that the individual has no occurrent or

standing desire for anything after retirement because the individual truly desires to die before she retires. For cases where we know for certain that a person does not have any occurrent or standing desire which saving for retirement would help fulfill, it does not seem clear that we would commonly say that it would be rational for the person to save for retirement.

It, therefore, seems to be at least two plausible ways in which The Desire-Based Interpretation can correspond to the fact that we commonly say things like *it would be rational for you to save for retirement even if you do not currently have any desires for any of the things saving for retirement helps achieve* and the fact that we commonly believe that statements like this can be true. Because of this, it does not seem as clear as we first might have thought that The Prudence-Based Interpretation is the most in line with this usage of 'rational action'. Based on these considerations, it might seem that we have reason to think that both of the interpretations are equally in accordance with this usage.

An argument from universal de facto authority in favor of The Desire-Based Interpretation

Statements about what action is rational, such as 'it would be rational for you to take your medicines', are normative statements about action. It appears that normally, when normative statements about future action are uttered, they are intended to influence the receiver to act in accordance with the statement. For example, if I say that *I think it would be rational for you to go to the local music festival*, it would normally be intended to influence you to go. Now, it seems that we commonly think that whenever we manage to convince someone to believe that a particular action is the only action which it is rational for them to perform, we highly expect them to be at least somewhat motivated to perform the action. This suggests that we think that rational action is such that whenever individuals believe that an action is the only action rational for them to perform, they will necessarily be at least somewhat motivated to perform the action. We can call this feature 'universal de facto authority'. It, therefore, seems reasonable to think that any understanding of 'rational action' is linguistically implausible if it implies that there are circumstances where an individual will not be the slightest motivated to perform an action she believes to be the only rational action for her to perform, or as we can say, implies that there are circumstances where rational action does not have any **de facto authority** on an individual.

Now, The Humean Theory of Motivation is the view that "belief is insufficient for motivation, which always requires, in addition to belief, the presence of a desire or cognitive state" (Rosati, 2016, p.11; see also Smith, 1987). This view is apparently the most prevailing among the alternative views on the matter (Rosati, 2016, p.11). If 1) individuals do not necessarily desire what is good for them, and 2) individuals only can be motivated to perform actions which they believe helps achieve something they desire, it seems to follow that I) it is not the case that an individual necessarily will have any motivation to perform an action she believes is the only one which will lead to the most good for her overall. Because of this, it seems that given The Humean Theory of Motivation, and given that an individual does not necessarily desire to do what would be good for her, it seems clear that The Prudence-Based Interpretation would imply that rational action does not necessarily always have any de facto authority. In other words, it seems to entail that it would be possible for there to be cases where an individual does not have any motivation to perform the only action she believes is rational for her to perform in the sense of The Prudence-Based Interpretation. This implication seems to suggest that The Prudence-Based Interpretation is linguistically implausible.

The Desire-Based Interpretation does not necessarily face this issue. Given The Humean Theory of Motivation, a version of The Desire-Based Interpretation will imply that what is rational necessarily always has de facto authority given that all the desires it proposes as rationally relevant goals are desires which necessarily will motivate an individual. Based on the considerations of this section, it, therefore, seems that out of the four interpretations discussed here, The Desire-Based Interpretation is the only interpretation we do not have reason to think is linguistically implausible. This, therefore, seems like the most promising approach to understanding rational action.

3.4 - The Holistic Interpretation or The Limited Role Interpretation?

In Section 3.1, I pointed out that one can choose between The Holistic Interpretation that The Instrumental Requirement is a requirement to perform the action which best helps fulfill the totality of the agent's rationally relevant desires, and The Limited Role Interpretation that

all The Instrumental Requirement provides basis for saying is what actions help fulfill what desires.

R. Jay Wallace discusses what he refers to as 'the holistic approach' without taking for granted The Desire-Based Interpretation. He describes the holistic approach as the view that "[p]ractical reason,[…] is a holistic enterprise, properly concerned not merely with identifying means to the realization of individual ends, but with the coordinated achievement of the totality of an agent's ends" (Wallace, 2018, p.14-15). This view is identical to The Holistic Interpretation apart from the fact that The Holistic Interpretation takes for granted that the rationally relevant desires of the actor is what should be regarded as what Wallace refers to as 'the ends of the agent'. Apparently, "many philosophers take [the] holistic approach to be the most promising way of thinking about the tasks of practical reason" (Wallace, 2018, p.15). With regard to how the holistic approach can be understood, Wallace writes that:

> The holistic approach finds its most sophisticated and influential expression in the maximizing conception of practical rationality. According to the maximizing conception, the fundamental task of practical reason is to determine which course of action would optimally advance the agent's complete set of ends. (Wallace, 2018, p.15)

An argument from linguistic plausibility favor of The Holistic Interpretation

As the following example illustrates, The Holistic Interpretation, i.e., The Holistic Approach given The Desire-Based Interpretation, seems very natural:

Example: The Gingerbread Dough Case

Imagine that an individual has a slight desire to eat a sizeable amount of gingerbread dough and has a strong desire not to experience the prolonged stomach pain she knows she inevitably will feel in the hours after doing so, and these are all the rationally relevant desires she has. Suppose that she has to choose between eating the dough and not eating it.

It seems that we would commonly think that it would be irrational for the individual to eat the gingerbread dough. This makes The Limited Role Interpretation seem linguistically implausible because it would not imply that it would be irrational for the individual to eat the gingerbread dough. The Holistic Interpretation, on the other hand, would imply that this is

irrational on the grounds that the alternative action, namely to not eat the dough, is an action that better helps fulfill the totality of her rationally relevant desires. On the face of it, this seems like a satisfying analysis of the case in terms of making sense of why we commonly think that it would be irrational for the individual to eat the dough. Because of this and because The Holistic Interpretation seems to have gained considerable acceptance, I am not going to argue further for why we should adopt this interpretation or why we should not adopt The Limited Role Interpretation. I will presume that an action only meets The Instrumental Requirement if it is the action that best helps fulfill the totality of the actor's rationally relevant desires.

Now, given that what The Instrumental Requirement requires an actor to intend to pursue is the best possible fulfillment of the totality of her rationally relevant desires, it seems reasonable to choose to call *the best possible fulfillment of the totality of an actor's rationally relevant desires, where she intends to achieve this goal* 'the rationally relevant goal'. Instead of saying that each rationally relevant desire which an actor intends is each a different rationally relevant goal, I will, therefore, say that the best possible fulfillment of the totality of the actor's rationally relevant desires is the rationally relevant goal of the actor, given that she intends to achieve this goal. I suspect this is the easier and more natural way to think about The Instrumental Requirement, and it makes parts of the discussions ahead easier to read. What Wallace calls 'the ends of the agent' is under The Desire-Based Interpretation ,the fulfillment of a single rationally relevant desire. Given The Desire-Based Interpretation, in cases where there is just a single rationally relevant desire, *the best possible fulfillment of the totality of the actor's rationally relevant goal* is identical to *the best possible fulfillment of this rationally relevant desire*. It does not seem to me like using this terminology to talk about The Instrumental Requirement changes the substance of the requirement in any way. However, it is important to keep in mind that this is the way I use the term 'rationally relevant goal'.

3.5 - Whether only intrinsic desires are rationally relevant desires

As mentioned in Section 3.1, one might ask whether instrumental desires, which are desires one has solely because achieving them helps achieve something else which one desires, can be rationally relevant desires, or whether only intrinsic desires, which are desires one does not have solely because their achievements are instrumental to achieving the fulfillment of some other desire which one has can be rationally relevant desires.

It seems reasonable to think that what is instrumentally desirable is not desirable in itself. It namely seems reasonable to think that the desirability of an instrumental desire is entirely dependent on it being instrumental in achieving an intrinsic desire. As Michael Smith writes: "It is a striking fact that instrumental desires disappear immediately an agent loses […] the relevant [intrinsic] desire […]. This is, if you like, part of what it is to be an instrumental desire, as opposed to [an intrinsic] desire" (Smith, 2004, p.96). Smith argues that we have an instrumental desire because we have 1) some intrinsic desire, and 2) a belief that achieving the instrumental desire helps fulfill the intrinsic desire (Smith, 2004, p.94). Consequently, Smith believes that instrumental desires are reducible to the intrinsic desire and means-end beliefs that explain them (Smith, 2004, p.96). Because the desirability of the fulfillment of instrumental desires seems to be derivative in this way of the desirability of the fulfillment of intrinsic desires, it seems reasonable to think that only the fulfillment of intrinsic desires is desirable in the relevant sense. Because of this, it seems reasonable to think that only intrinsic desires can be rationally relevant desires. Without diving deeper into this topic, I will, therefore, presuppose that The Intrinsic Interpretation is correct and that only intrinsic desires can be relevant desires with regard to The Instrumental Requirement.

3.6 - Whether criticism is relevant

In Section 3.1, I mentioned The Relevance of Criticism Interpretation which is the view that an intrinsic desire can be rationally irrelevant due to the fact that it can be criticized with reference to other intrinsic desires, and The Irrelevance of Criticism Interpretation which is the view that an intrinsic desire cannot be rationally irrelevant because of this.

I mentioned that one version of The Relevance of Criticism Interpretation is the view that an intrinsic desire is rationally irrelevant if it can be criticized on the basis that it is conflicting with another desire that is "subjectively more important to the agent" (Wallace, 2018, p.14).

Two intrinsic desires of an actor are in conflict if they suggest that the actor performs different actions. Now, it seems that virtually any two intrinsic desires will have some contradictory implications for what actions they suggest an actor perform. Take two intrinsic desires, such as *the intrinsic desire to provide as best as one can for one's family* and *the intrinsic desire to complete a big hobby project as quickly as possible*. It seems clear that the action that best helps fulfill the first intrinsic desire would typically not always be the same action that best helps fulfill the second. This would apparently imply that virtually all desires will be criticizable in this way except the one that is the most superordinate. The view that only an intrinsic desire that is regarded as being the most superordinate is a rationally relevant desire seems linguistically implausible. It namely seems that if an actor only can perform two actions and one action fulfills all the intrinsic desires an actor has completely apart from the most superordinate desire, while the alternative action only fulfills the most superordinate desire to a small degree, then we would commonly say that it would be rationally required for the actor to perform the first action. The principle that *an individual is rationally required to hold intrinsic desires that do not conflict with the intrinsic desires she deems to be the most important ones*, therefore, seems to be linguistically implausible

Now, a different version of The Relevance of Criticism Interpretation is the view that 1) an intrinsic desire can be criticized on the basis that a given actor having the desire leads to less overall fulfillment of her intrinsic desires compared to not having the desire because having the desire overall hinders the actor from performing actions which are rational for her to perform and that 2) an intrinsic desire is rationally irrelevant if this criticism applies. Consider the following example to see what this view would imply:

Example: The Alcohol Case

Suppose that an individual has a strong intrinsic desire to drink alcohol and that because of that she often chooses to drink alcohol in situations where that is not the action which best helps fulfill the totality of her rationally relevant desires, and, therefore, is rationally required in those situations not to drink alcohol. Imagine that if she did not have such an intrinsic desire to drink alcohol, this would lead to her rationally relevant intrinsic desires being better fulfilled compared to her having the desire because she would more often be able to choose to do what is rational for her to do.

In such a scenario, the person's intrinsic desire to drink alcohol would be a hindrance to her ability to do what best helps fulfill her rationally relevant desires. If she had the choice between *ridding herself of that intrinsic desire, which would lead to more overall fulfillment of the rationally relevant intrinsic desires she has*, and *not ridding herself of it*, all other things being equal between the options, *ridding herself of the desire* would be the action which would best help fulfill her rationally relevant desires overall. It, therefore, seems clear that one can criticize an intrinsic desire based on The Instrumental Requirement, in the sense that not having it would be better for fulfilling the totality of the actor's rationally relevant intrinsic desires.

Statements such as *one would be rationally required to rid oneself of an intrinsic desire to shop luxury items, all things being equal, if having this intrinsic desire overall hinders one's ability to do what best helps fulfill one's rationally relevant intrinsic desires overall* do seem very intuitive. If it is correct that intrinsic desires can be criticized in this way, it would seemingly provide the bases for criticism of many intrinsic desires, as it is conceivable that many different intrinsic desires overall hinders an individual having one from doing what best helps fulfill her rationally relevant intrinsic desires.

However, even given that it is correct that intrinsic desires can be criticized in this way it seems that we should not think that only intrinsic desires that cannot be criticized in this way are rationally relevant desires. The reason for this is that this criticism is based on the claim that *it is rationally required to rid oneself of a given desire given a chance and all other things being equal*. From this, it simply does not follow that such a desire is not a rationally relevant desire, nor does it seem to give us a reason to suppose this.

It, therefore, does not seem that we have reason to think that intrinsic desires can be criticized with reference to other intrinsic desires in a way that makes them rationally irrelevant. Because of this, it seems that we have reason to adopt The Irrelevance of Criticism Interpretation.

3.7 - Whether the future should be included

As described in Section 3.1, The Contemporary Desires Interpretation is the view that only desires one has when one is making a choice between actions can be rationally relevant desires, while The Future Inclusive Interpretation is the view that in addition to these desires all the desires which the actor will or is anticipated to have at some point in the future can be rationally relevant.

Arguments in favor of The Future Inclusive Interpretation

Those favoring The Future Inclusive Interpretation might argue that it would be arbitrary to say that only one's contemporary desires and not the desires one will have in the future but which one is not aware of yet are relevant desires with regard to The Instrumental Requirement. However, if we think that an essential feature of a concept of rational action is that it has universal de facto authority, this can be used to justify the view that only one's contemporary desires are rationally relevant desires, as I will shortly show. This would prevent the view that future desires cannot be rationally relevant desires from being arbitrary.

Another argument in favor of The Future Inclusive Interpretation, which might seem tempting, is the following: 1) *Fulfillment of either type of desire will lead to desire satisfaction*. 2) *Desire satisfaction is what is good for an individual*. 3) *An individual is rationally required to do what leads to the most overall good for herself*. I) Therefore, an individual is rationally required to do is to do what leads to the most satisfaction of *all* of her desires, including both contemporary and future desires. However, the third claim is the claim of The Prudence-Based Interpretation, and it is a central premise of this argument. The Prudence-Based Interpretation, therefore, needs to be established in order for the argument to be convincing. I have already argued why The Desire-Based Interpretation seems more linguistically plausible than The Prudence-Based Interpretation. Without further argument for The Prudence-Based Interpretation, the above argument, therefore, does not seem convincing.

An argument from universal de facto authority in favor of The Contemporary Desire Interpretation

It seems reasonable to think that if an individual has a contemporary intrinsic desire, this will usually cause the individual to have at least some degree of motivation to perform an action if she believes it will help fulfill that desire. Furthermore, it seems reasonable to think that under all normal circumstances, this is the *only* way a motivation to perform an action can be created in a person. In other words, that under all normal circumstances, an individual will only have a motivation to perform a given action if she has a contemporary desire that she believes the given action can help her fulfill. This would mean that under all normal circumstances, the only way future desires can cause an individual to be motivated to perform an action, is if the individual has a *contemporary* desire to help fulfill these future desires. However, as an individual does not necessarily have a contemporary desire to help fulfill her future desires, it seems clear that there could be cases where an individual would not be motivated to perform what she believes to be rational in the sense of The Future Inclusive Interpretation. The following is an example of this:

Example: The Children in Ten Years Case

Say an individual has no contemporary desire to have children in ten years, but that she knows for a fact that she in ten years will have a great desire to have kids. Let us suppose, for the sake of illustration, that she will have no other future desires and has no other contemporary desires. Suppose that there is a pill in front of her, which she can take for free, which will make her permanently sterile and create no other side effects, and that she has the choice between taking and not taking the pill.

In such a scenario, the implications of The Future Inclusive Interpretation and The Contemporary Desires Interpretation for what action it is rationally required for the individual to perform would differ. The All Future Desires Implication would, namely, imply that it would be irrational for her to take the pill, while The Contemporary Desires Interpretation would not. If the claims of the discussion above are correct, it is the case that unless she has a contemporary desire to have this future desire fulfilled, she will not be motivated to act on the basis of fulfilling this future desire. In the scenario above, where she has no such contemporary desire, The Future Inclusive Interpretation would, therefore, imply

that the action it is rationally required for her perform is an action she cannot be motivated to perform.

It, therefore, seems clear that The Future Inclusive Interpretation is an interpretation which would entail that it is not necessarily always the case that an individual would have any motivation to perform an action she believes to be the only action which it is rational for her to perform in the sense of The Future Inclusive Interpretation. In other words, that it does not have universal de facto authority. In Section 3.3, I pointed out that an understanding which entails that rational action does not have universal de facto authority seems linguistically implausible.

Based on the considerations of this section, it, therefore, seems that we have reason to think that The Contemporary Desires Interpretation Is the most linguistically plausible and that this is the interpretation we should adopt.

3.8 - What to take contemporary intrinsic desires to be

In Chapter 3, I explained that one can choose between The Standing Inclusive Interpretation, which is the view that both occurrent desires and standing desires can be rationally relevant desires, and The Occurrent Desires Interpretation, which is the view that only occurrent desires can be rationally relevant.

An argument from linguistic plausibility in favor of The Standing Inclusive Interpretation

An objection that Korsgaard has against the view that The Instrumental Requirement is desire-based is that she thinks that it entails that it is impossible to act irrationally (Korsgaard, 1997, p.40). Such an implication would be a serious problem for any understanding of rational action, because it is clear that we commonly think that humans sometimes perform irrational actions. That the desire-based interpretation entails that this is impossible, would, therefore, seem to make it linguistically implausible. It, therefore, seems that we should regard the following as a requirement for any understanding of rational action:

The Fallible Requirement

It is a requirement of a concept of rational action that it does not entail that it is impossible for humans to act irrationally.

This requirement also seems to be implied by what Douglas Lavin calls 'the error constraint', which he formulates as the following: "*an agent is subject to a principle only if the agent can go wrong in respect of it*" (Lavin, 2004, p.425). Lavin, plausibly enough, takes to be a general requirement of normative principles.

Now, it seems likely that The Occurrent Desires Interpretation conceivably entails that it is impossible to act irrationally. It, namely, seems plausible that when individuals act voluntarily, the only occurrent desire they have at the exact moment they initiate the given action is to perform the action, and when doing so, they intend to fulfill this desire. It seems plausible, for example, that if I voluntarily kick a ball, the only occurrent desire I have at the exact moment I decided to do so was to kick the ball and that I had the intention to fulfill this desire when kicking the ball. What it is rationally required for an actor to do according to The Occurrent Desire Interpretation is to perform the action which best helps fulfill the contemporary intrinsic desires that are present in his psyche at the moment. I might have had another occurrent desire just a moment before, but it seems that such a previously occurrent desire should not be regarded as a rationally relevant desire if we take The Occurrent Desires Interpretation seriously. If one thinks that desires which were occurrent just moments before one performs a given action can be rationally relevant desires, it namely seems that it would be very hard to explain why not other standing desires can be rationally relevant. To avoid the charge that The Occurrent Desire Interpretation contains the seemingly arbitrary claim that only some standing desires can be rationally relevant, it, therefore, seems that given The Occurrent Desire Interpretation, one would have to maintain that only the occurrent intrinsic desires at the exact moment of decision can be rationally relevant.

However, if 1) the view of The Occurrent Desire Interpretation is that what it is rational for an actor to do is to intentionally do what best helps fulfill her occurrent intrinsic desires at the moment of the action, and 2) it is true that the only occurrent desire an individual has at the exact time she voluntarily initiates an action is to perform the action, and an individual always intends to fulfill this desire when performing an action voluntarily, this seems to entail that one is always acting in accordance with what is rational when one is acting voluntarily.

Given this, The Occurrent Desire Interpretation would violate The Fallibility Requirement and, therefore, seem linguistically implausible.

The Standing Inclusive Interpretation, however, does not seem to entail that it is impossible for an individual to act irrationally. This is because under this interpretation what action is rational is not merely determined by the occurrent desires of an individual, but also by her standing desires, and it seems clear that an individual will not necessarily perform the action which best helps fulfill the totality of her occurrent and standing desires when acting voluntarily. For illustration, consider the following case:

Example: The Election Case

Say it is an important election day. Imagine that an individual has a strong standing desire to vote in this election, but that she has forgotten that that day is her last chance to vote and, therefore, does not have an occurrent desire to go and vote. Suppose also that the only occurrent desire the individual has all day is to watch a tv show, but that this occurrent desire is weaker than her standing desire to vote in the election.

It seems clear that it in such a scenario would be entirely possible for the individual to not act on her standing desire to vote in the election. Given that these are all the rationally relevant desires the individual has, The Standing Inclusive Interpretation implies that it would be rational for the individual to go and vote. Hence, it seems clear that this interpretation meets The Fallibility Requirement.

A second argument from linguistic plausibility in favor of The Standing Inclusive Interpretation

Furthermore, The Standing Inclusive Interpretation seems to be more in accordance with what we think are rational actions than The Occurrent Desires Interpretation is, as the following case seems to show:

Example: The Completing Education Case

Let us say that an individual who is having a difficult time completing her education, generally for a long time has had, and continues to have, a great standing intrinsic desire to complete her education, and that she tomorrow morning has an exam she needs to do well at in order to achieve her goal. Let us say that in the moment, because she is intoxicated, this standing intrinsic desire is not on her mind,

and the only occurrent intrinsic desire she has is to go to a party with her friends. Acting on this desire, however, would ruin her chances of doing well on the exam.

It seems that we would not commonly suggest that, given that these are all the desires the individual of the case has, what it is rationally required for her to do is to go to the party, even though this is her only occurrent desire in the moment, and that we, in fact, would commonly claim that it would be irrational for her to go to the party with her friends. The Occurrent Desires Interpretation would imply that it is rational for the individual to go to the party, while The Standing Inclusive Interpretation would imply that it is not. This suggests that the latter interpretation is more linguistically plausible than the former.

Given what has been considered in this section, it seems that we have reason to suppose that The Standing Inclusive Interpretation is the interpretation we should adopt.

3.9 - Whether only laundered desires can be rationally relevant

In Section 3.1, I mentioned that one can hold the view that only desires that have met some requirement of having been formed in a certain way are rationally relevant desires. I mentioned that The Deception Laundering Interpretation and The Mental State Laundering Interpretation are respectively the views that desires that are formed on the basis of false beliefs and formed due to an individual being in certain temporary states of mind are not rationally relevant desires, and that The No Laundering Interpretation is the view that there are no laundering requirements a desire needs to meet in order to be a rationally relevant desire.

Questioning the reasonableness of The Deception Laundering Interpretation

It seems that we have reason to doubt the claim that intrinsic desires that are based on false beliefs are not rationally relevant desires. To start exploring why let us consider the following case:

Example: The Athlete Case - Version 1

Imagine a scenario where a sports fan has the apparently intrinsic desire that *a given athlete wins the 10 000 meter run event in the Olympic Games*. Suppose that this apparently intrinsic desire was formed only on the basis of her belief that the athlete is a Christian, and that this belief was formed on the basis of the fact that the athlete always is wearing a necklace with a cross. Now suppose that it is revealed to the sports fan that the athlete is not a Christian after all and was wearing the necklace in honor of her deceased friend. Suppose that the sports fan, therefore, stops being a fan of the athlete and stops having the apparently intrinsic desire that the athlete wins the event.

It does seem natural to suppose that the sports fan's desire that the athlete wins is not a rationally relevant desire. One might try to explain this by saying that this desire is not a rationally relevant desire because it was formed based on a false belief, thereby suggesting that there is some sort of deception laundering requirement. However, it might seem reasonable to think that the desire is not really an intrinsic desire at all. Given that the desire is formed only on the basis of the belief that the athlete is a Christian, it namely seems reasonable to instead interpret the case as the fan having an intrinsic desire *to see some Christian athlete win the event*. Given this interpretation, one could explain that the fan's desire is not a rationally relevant desire because it is not truly an intrinsic desire. This would mean that a deception laundering requirement is not needed to explain why the sports fan's desire is not rationally relevant.

Now, it seems plausible that false beliefs *can* lead to an individual developing an intrinsic desire. Consider another version of The Athlete Case:

Example: The Athlete Case - Version 2

Imagine a scenario where everything is the same as in The Athlete Case - Version 1, except that when the sports fan learns that the athlete is not a Christian, the sports fan does not stop having the desire that the athlete wins the Olympic event, despite the fact that the desire was originally formed solely on the basis of the belief that the athlete is a Christian. Let us suppose that the desire has turned into a truly intrinsic desire of the sports fan simply due to the fact that she has had the desire for a longer period of time.

This example seems plausible, and it, therefore, seems plausible that a false belief can lead to an individual having an intrinsic desire in this way. If the desire of version 2 is not a rationally relevant desire, we cannot explain that by saying that it is not truly an intrinsic desire, as it is stipulated in the example that it is. However, it might not be easy to see why an

intrinsic desire formed in this way would not be a rationally relevant desire. In version 1 of the case, the sports fan stops having the desire when she learns the truth. It seems to me like this fact is the main reason why it seems natural to suppose that the desire of version 1 is not a rationally relevant desire. However, in version 2, the sports fan does not stop having the desire when she learns the truth. It is not that she is ignoring the fact that the athlete is not a Christian. She recognizes that fact but does not stop having the desire for the athlete to win. It does not seem clear at all to me that such a desire would not be a rationally relevant desire.

Furthermore, it seems clear that we commonly have intrinsic desires which are not based on any beliefs, true or false. For example, it seems that we commonly have an intrinsic desire for the pain to stop if we are in pain. It seems reasonable to think that such a desire is usually an intrinsic desire we are born with and that it, therefore, is not formed on the basis of true beliefs about the world. It seems clear that an inborn intrinsic desire for the pain to stop would be a rationally relevant desire. This would mean that there is no requirement to be a rationally relevant desire that it is formed only on the basis of true beliefs. Given this, it seems reasonable to suppose that intrinsic desires formed on the basis of false beliefs can be rationally relevant desire too. Based on what has been considered here, it seems that we are not justified in making the assumption that desires can only be rationally relevant desires if they are not formed on the basis of false beliefs. In fact, based on these considerations, it seems more reasonable to think that there is no requirement in order for a desire to be a rationally relevant desire that it is not formed on the basis of false beliefs.

An argument from superfluity against The Mental State Laundering Interpretation

It seems natural to think that emotional states, such as rage, often make people act irrationally. Strong emotions can clearly cause people to have occurrent intrinsic desires which they would not have had, had they not been in that emotional state. For example, it seems clear that rage can cause one to have the occurrent intrinsic desire to break the things around oneself. It seems natural to think that throwing an item one normally does not want to break, to the ground, such as a vase one likes having due to rage, would often be irrational. It might, therefore, seem natural to dismiss intrinsic desires caused by a mental state such as rage as being irrelevant with regard to The Instrumental Requirement. This observation may

lead us to think that an individual's intrinsic desires are only rationally relevant if they are produced by certain mental states of the individual.

Now, everything considered so far in this thesis seem to suggest that an action is irrational for an individual to perform if it is not the action that would best help fulfill the totality of all of her contemporary intrinsic desires. For the purposes of the rest of this section, I will refer to this as Theory R. Given Theory R, one might question the need for a mental state laundering requirement to explain how it can be that actions performed while being in an emotional state are often irrational. Let us grant that it is often irrational for individuals to break something while being angry. Given that we think that standing desires in addition to occurrent desires can be rationally relevant desires we might explain this by saying that in many such cases *not breaking the given item* better helps fulfill the totality of all of the individual's contemporary intrinsic desires compared to *breaking the given item*, and this is what makes it so that breaking something while being angry often is irrational.

I will shortly test the notion that one can explain all clear instances of irrational action caused by some non-ordinary mental state under Theory R. However, before I do a short discussion regarding how the strength of contemporary intrinsic desires might be understood under Theory R is necessary.

It namely seems reasonable to suppose that an individual's intrinsic desires can be of different strengths in the sense that one can desire something, such as *becoming a mother*, more than one desires something else, such as *one's favorite team winning the world cup*. Given this, one might wonder how this type of strength should be defined and how the absolute strength of an intrinsic desire is to be determined. However, for the purposes of this thesis, I will content myself with the following understanding for which of two intrinsic desires has the most of this type of strength:

The Ideal Conditions Understanding

The strongest of two of an actor's intrinsic desires is the one the actor would choose to fulfill if she was made completely aware of the two desires, had been considering them both carefully, could fulfill one and only one of them completely, and all other things were equal between the options.

To use the term 'stronger desire' to describe a desire which is stronger than another in this way might be problematic because the term normally means something different and,

therefore, the connotations it has seem likely to interferes with our intuitions when this type of strength is discussed. In order to avoid this, I will, therefore, instead say that the desire an individual would choose to fulfill over another under these conditions is **more weighty** than the other desire. It seems reasonable to suppose that all intrinsic desires have some absolute strength of this sort, or as I will say **weightiness**, though the absolute weightiness of an intrinsic desire might be significantly more difficult to determine than which of two intrinsic desires is more weighty. From now on, when I use adjectives such as 'slight' or 'strong' to describe a desire, I mean to refer specifically to the weightiness of the desire.

Now I will return to the task of testing the notion that one under Theory R can explain all clear instances of irrational action caused by some non-ordinary mental state without appealing to a laundering requirement. One might think that the following example is a counterexample to this notion:

Example: The Wall Case

Suppose that an individual has a slight standing intrinsic desire to not have a small bruise on her hand. Imagine, however, that one morning she is angry, she has a more weighty occurrent desire to hit the wall with her hand and does so, getting a small bruise on her hand, but experiencing no pain. Suppose that right after she stops being angry, she becomes aware of her standing desire, and because of that, it turns into a slight occurrent desire to not have the small bruise on her hand. Let us suppose that she has no other rationally relevant desires.

The reason this might seem like a counterexample to the notion above is that Theory R alone would not imply that it is irrational for the individual to hit the wall with her hand. If it seems clear that we would commonly insist that it is, it would give us reason to think that the theory cannot explain these cases and that we need a type of mental state laundering requirement. However, it is not so clear that for the individual to hit the wall in the example above is irrational. After all, her desire to hit the wall is more weighty, meaning that if she hypothetically had been completely aware of both desires at the same time and considered them carefully, she would have chosen to fulfill her desire to hit the wall. One might try to respond by claiming that an individual in such a scenario would likely regret hitting the wall after the fact. Let us suppose that the individual does regret hitting the wall in the example above. Now, the individual might have a contemporary intrinsic desire not to feel regret, or

she might not have such desire. If she does not have such a contemporary intrinsic desire, and there still are no other rationally relevant desires than the ones mentioned in the original case, it does not seem clear how the fact that she regrets hitting the wall would make the act irrational. If she, on the other hand, does have a contemporary intrinsic desire not to feel regret this would change the example somewhat. Under Theory R, this would be another rationally relevant desire which one must take into consideration. Given that there are no other rationally relevant desires than the three desires mentioned, in order for Theory R to imply that it is irrational for the individual to hit the wall in this new version of the case, the individual's desire to hit the wall would have to be more weighty than both her desire to not have a small bruise on her hand *and* her desire to not feel regret combined. If the intrinsic desire to hit the wall was that strong compared to the two other rationally relevant desires, then it again does not seem clear that hitting the wall would be irrational. Therefore, it does not seem clear that The Wall Case is a successful counterexample to the notion that one can explain all cases where an individual acts irrationally due to a non-ordinary mental state by saying that her actions are not what would best help fulfill the totality of her contemporary intrinsic desires.

The well-known principle of Occam's Razor is the principle that a theory regarding a phenomenon should make as few assumptions about the world as possible while being able to explain the phenomenon. According to the principle of Occam's Razor, we should prefer a theory of rational action which can explain our use of the term 'rational action' without making the extra assumption that only desires produced by certain mental states are rationally relevant desires compared to one that does, all other things being equal. Given these considerations, it, therefore, seems that we have reason not to make the assumption that only desires that are produced by certain mental states are rationally relevant desires.

In lack of compelling reason to think that we should think that there is some laundering requirement which desires need to meet in order to be rationally relevant desires, it seems that we have reason to think that we should not discriminate between desires on this basis and that we have reason to adopt The No Laundering Interpretation.

3.10 - Belief, evidence, or actuality?

As described in Section 3.1, The Objective Interpretation is the view that *The Instrumental Requirement requires is that the given actor performs the action which actually will best help fulfill her rationally relevant desires*, The Subjective Interpretation is the view that *what it requires is that the given actor performs the action which she believes is the action which will best help fulfill her rationally relevant desires*, and The Evidence-Centered Interpretation is the view *that what the requirement requires the given actor to do is to perform the action which the evidence held by the actor suggests is the action which best will help fulfill the rationally relevant desires of the actor*. I will take *the evidence held by the actor* to be identical to *all the information the actor is aware of*.

An argument from linguistic plausibility against The Objective Interpretation

The Objective Interpretation has an implication which makes it seem linguistically implausible. The Objective Interpretation, namely, seems to entail that an individual not always will be at fault for performing an irrational action (Wallace, 2018, p.19). An individual can fail to do what is rationally required for her to do in the sense of The Objective Interpretation, for at least three different (explanatory) reasons: A) not do what she correctly believes is rationally required of her, B) not having the correct belief about what action is rationally required, when she has reason to believe it, and C) not having information which would give her a reason to adopt the correct belief about what action is rationally required for her to perform.

An example of the first would be if I correctly believe that it, in a certain situation, is rationally required of me to not go into a war zone, but I do it anyway. An example of the second would be if I have reason to believe it is rationally required of me to not go into the war zone, but I do not adopt this belief, and because of the lack of this belief go into the war zone. An example of the third would be if I did not adopt the correct belief that it is rationally required of me to not go into the war zone, because the information I have gives me no reason to adopt this belief. I might, for example, not have heard the news that a war has broken out in that area.

While it seems at least conceivable that an individual is always at fault in some sense, for failing to do what is rationally required for one of the first two reasons, it seems clear that an individual not necessarily will be at fault for failing in the third way. An individual might have done what can reasonably be expected of her to gather such information, but still not obtained information which would have given her reason to adopt the correct belief about what is rational for her to do. This might make it so that she does not do what, according to The Objective Interpretation, is rationally required of her to do, even though she believed everything she had reason to believe and acted in accordance with what she believed was rational for her to do. In such a scenario, it would not seem like a linguistically plausible use of the term 'at fault' to say that the individual is at fault for failing to do what, according to The Objective Interpretation, is rationally required of her to do. The Objective Interpretation thereby seemingly entails that one is not necessarily at fault when one is performing an irrational action, and this suggests that it is linguistically implausible.

Furthermore, it seems that we do not commonly say that an action is irrational despite not being the action which actually best helps fulfill the rationally relevant desires of the actor unless the actor either believes or has reason to believe that her action is not the action which best helps fulfill her rationally relevant desires. The following example seems to illustrate this:

Example: The Bank Case

Imagine that an individual wants to store her money as securely as possible for a year, and she has the option between putting the money in a bank or keep it in her car parked on the street. Suppose that she believes and has every reason to believe that if she puts the money in the bank, she has one in a million chance of not getting it back in a year and that if she keeps the money in the car parked on the street, she has a much lower chance of not having it in a year because it is likely that her car will be broken into during that time. Suppose that she, therefore, puts the money in the bank, believing that the money will be more secure there. Now, imagine that the year passes and that some time during the year, unlikely events cause the bank to go bankrupt and that she, therefore, does not get her money back. Imagine also that her car has not been broken into that entire year and that she would still have had her money had she put the money in her car instead of the bank.

Given that the individual's only rationally relevant desire was to have access to the money after one year The Objective Interpretation would imply that it was irrational for the

individual to put the money in the bank instead of the car because putting the money in the car would have been the action which actually would have best helped fulfill this desire. However, it seems clear that we would not commonly say that the rational thing for the individual to do would have been to put the money in the car or that putting the money in the bank was irrational. This, despite the fact that putting the money in the car was the action that actually would have best helped the fulfillment of the rationally relevant desires of the actor. This seems to clearly suggest that The Objective Interpretation is linguistically implausible.

It seems natural to think that the reason why we commonly would deny that it was irrational for the actor to put the money in the bank instead of the car in the case above, is because she believed or because she had every reason to believe the putting the money in the bank was the more secure option. This suggests that The Subjective Interpretation and The Evidence-Centered Interpretation are more in line with this use of the term 'rational action' than The Objective Interpretation is.

An argument from linguistic plausibility against The Subjective Interpretation

The Subjective Interpretation would imply that an action is rationally required for an individual to perform if the individual believes that the action is the that which will best help fulfill her rationally relevant desires. However, there seem to be cases that suggest that this is linguistically implausible. Consider the following example:

Example: The Vaccination Case

Imagine that a father's only rationally relevant desire is for as little harm as possible to fall on his child. Suppose that the information he is aware of clearly suggests that not vaccinating his child would be much more likely to cause the child more harm than vaccinating his child would. However, imagine that because he does not realize what the information implies he believes that vaccinating his child would be more likely to cause more harm to his child than not vaccinating his child would and that he because of this with the intention to minimize the amount of harm to his child prevents his child from getting vaccinated.

It seems clear that we would commonly say that for the father to prevent his child from being vaccinated under such circumstances would be an irrational action. The Subjective Interpretation would imply that performing this action would be rationally required of the father. This makes this interpretation seem linguistically implausible.

Now, one might try replying that the father only would be irrational in a theoretical sense, meaning that he fails to adopt a belief which he would be rationally required to adopt based on his other beliefs and that his action is, therefore, not irrational. In other words, one might claim that an actor being theoretically irrational cannot cause his actions to be irrational. However, while it seems correct that the individual is irrational in the theoretical sense and that this is what is causing the father to act the way he does, as long as it seems linguistically implausible to say that the father's action of preventing his child from getting vaccinated would not be irrational, it seems that we should adopt the view that his action *is* irrational and that this is due to the fact that he fails to be theoretically rational, thereby denying that actions cannot be irrational simply due to the actor being theoretically irrational.

The Evidence-Centered Interpretation would imply that it would be irrational for the father to prevent his child from getting vaccinated because the information the father is aware of suggests that having his child vaccinated would be the action which best will help fulfill his only rationally relevant desire of minimizing the harm to the child. It, therefore, seems that we have clear reason to think that out of the three interpretations considered here, The Evidence-Centered Interpretation is the most linguistically plausible and is the one we should adopt.

3.11 - Wide-scope or narrow-scope?

In Section 3.1, I mentioned The Narrow-Scope Interpretation which suggests that a person can only meet The Instrumental Requirement by either performing the actions which best help fulfill the relevant goal one happens to have and The Wide-Scope Interpretation which suggests that one in addition to this can meet the requirement by changing her rationally relevant goal (Wallace, 2018, p.12-13).

An argument from linguistic plausibility against The Narrow Scope Interpretation

Jonathan Way writes the following:

> The Wide-Scope view is typically motivated by a problem for the Narrow-Scope view. Narrow-Scope requirements are susceptible to clear counterexamples. We can construct such

> examples by considering cases in which one of the antecedent attitudes is itself irrational. For example, consider a case […] in which you intend to do something which is obviously crazy. In such a case, it is implausible that rationality requires you to have the consequent attitude—on the contrary, the consequent attitude would itself be irrational. (Way, 2011, p. 228)

Here Way mentions the example of when a goal is obviously crazy, which he believes to be a counterexample to The Narrow-Scope Interpretation and the idea that it is never rationally required for an individual to change her rationally relevant goal. Now, Way seems to assume that the relevant goal of The Instrumental Requirement is simply what an individual intends. This seems clear when he formulates the requirement as the following: "If you intend to E and believe that M is necessary for E, then rationality requires that you intend to M" (Way, 2011, p. 228; see also Rippon, 2011). This entails The Actual Intention Interpretation discussed in Section 3.1 and Section 3.3. Now, The Narrow-Scope does not seem vulnerable to this kind of counterexample under The Desire-Based Interpretation.

Under The Desire-Based Interpretation it seems conceivable that the rationally relevant goal of an individual can be crazy in some sense of the word. For instance, say that an individual's only rationally relevant desire is to go to prison. Such a desire might conceivably be crazy in some sense of the word, and because the rationally relevant goal in such a scenario would be the fulfillment of this single desire, it seems that this goal would be crazy in the same sense. However, it does not seem clear that we would commonly say that it would be irrational for this individual to perform the action which her evidence suggests best will help fulfill this desire if we truly believed this to be her only rationally relevant desire. The example of a crazy rationally relevant goal, therefore, does not seem to work as a clear counterexample under The Desire-Based Interpretation.

An argument from The Rational Coherence View

Now, those favoring The Wide-Scope Interpretation might argue that any requirement of rationality, including both any requirement of theoretical rationality and any requirement of practical rationality, is a requirement to hold a coherent set of attitudes, such as beliefs, desires, or intentions. The requirement of theoretical rationality, one might contend, is to have beliefs that are coherent with one another (Wallace, 2018, pp.12-13). In the case of instrumental rationality, one might claim, one is rationally required to hold a coherent set of

intended ends, beliefs about what means will help achieve what ends, and intended means (Kolodny & Brunero, 2018, pp.10-35). We can call this view 'The Rational Coherence View'. One might argue based on such an assessment, that it would be arbitrary to hold that The Instrumental Requirement is a requirement that one's intended means and goal are made coherent by means of changing one's intended means, because one can obtain such a coherency by means of changing one's goal too. However, that The Instrumental Requirement is a requirement to be coherent in this way is not obvious. Niko Kolodny argues, for instance, for the view that the true requirements of reason is "a requirement to believe what the evidence supports, [and] intend what promises to be worthwhile" (Kolodny, 2008, p.462). Because the view is not obvious, The Rational Coherence View, therefore, requires support. We should, therefore, consider whether we seem to have reason to think that The Rational Coherence View is correct.

An argument from explanation in favor of The Rational Coherence View

It might seem natural to argue that the view that The Instrumental Requirement is about rational coherence makes us able to make sense of why it is rational to perform the actions which the evidence one holds suggest will best help one achieve one's rationally relevant goal. One might explain that this is rational because we are rationally required to be rationally coherent, and if one intends to achieve such a goal, one must intend such means in order to be rationally coherent. However, this is no satisfying explanation in itself because it seems appropriate to ask why we are rationally required to be rationally coherent. One might try replying that what we typically mean to say when we call an action rational is that *intending to perform the action is necessary for the actor to be rationally coherent given the goal she intends to achieve*, but the problem is that it does not seem clear why we should not think that when we call an action rational we typically simply mean that say that *the action is the one which the evidence one holds suggest will best help fulfill one's rationally relevant goal*. This reply, without any further justification, is, therefore, to beg the question. In order to be an explanation we should think is correct, an explanation needs to point to claims we have independent reason to believe are true, not just move the question from one unjustified claim to another. It, therefore, seems clear that reasons for why we should think that one is rationally required to be rationally coherent is needed.

An argument from explanatory power in favor of The Rational Coherence View

It seems that we commonly think that the two following statements are correct:

*All other things being equal, one *should* do what leads to something which is valuable*,
and
*We *should* act in accordance with what is rationally required, all other things being equal*.

It might seem that one reason for why one might think that The Rational Coherence View is correct is that one can connect the two instances of 'should' and thereby demystify why we commonly think that both of the above statements are correct. One could explain the second statement by saying that 1) one should do what leads to what is valuable, all other things being equal, that 2) it is valuable to have one's mental entities such as beliefs, desires, and intentions coherent with one another,, therefore, I) "[being rationally coherent] is among the things that agents […] ought, to do or intend […]." all other things being equal, "just as we ought not to torture, or ought to care for our children, we ought to be rationally coherent", (Kolodny & Brunero, 2018, p.5), and because of this we should act in accordance with what is rationally required all other things being equal. This would mean that one could explain that we think both of the statements are true because we think one should do what leads to something which is valuable, all other things being equal.

However, the argument would only work if the 'should' of *one should always do what leads to something which is valuable* and the 'should' of *one should always do what is rationally required* has the same meaning. That this is the case is something one might seriously question. It might seem reasonable to think that 'should' when used in a similar way as in the first statement, usually is meant as 'should' in the sense that one should bring about what is valuable in and of itself, regardless of whether or not it would be in the interest of the individual herself or whether she desires it. 'Should' when it is used in a similar way to how it is used in the second statement, however, might usually be meant as 'should' in either the sense that one should do what is leading to the best outcome for oneself or in the sense that one should do what is leading to the most fulfillment of one's desires. If the 'should' of the two statements are meant in different senses, the conclusion will not follow from the premises, and the deduction above would be invalid and unsound. If they are meant to be the

same sense, however, one of the premises might seem implausible, which would also be leaving the deduction unsound.

A second argument from explanatory power in favor of The Rational Coherence Interpretation

It might be desirable to explain why we commonly say that *it is irrational to not believe what one has all things considered reason to believe*, and that *it is irrational not to perform one of the actions which the evidence one holds suggest will best help one achieve one's rationally relevant goal*. One can hold the view that the first statement would be implied by what we can refer to as 'The Requirement of Theoretical Rationality' and understand this as a requirement to have coherence between one's beliefs. The second statement is implied by The Instrumental Requirement, and one can hold the view that this is a requirement to have coherence between one's rationally relevant goal and intentional actions. Accordingly, one can think that both The Instrumental Requirement and The Requirement of Theoretical Rationality are requirements to be rationally coherent. Another argument for The Rational Coherence View, therefore, can be that it makes it possible to explain why we call both of these things in the statements above 'irrational'. Given the two views, one can, namely, explain that what makes something irrational is lack of rational coherence, and since both of these are instances where an individual is not rationally coherent, there is no wonder why we say that both of these are instances of irrationality. However, there are at least a couple of factors that prevent this argument from being convincing.

Firstly, one might think that the fact that we call both of these things 'irrational' does not need an explanation, any more than the fact that we use the word 'date' to refer to both a day of the year and a meeting with a love interest. We use the same words for many different concepts where the different uses of the words are not applied according to a shared principle.

Secondly, the two instances might be connected in another way. If it is the case that one is rationally required to adopt the means which the evidence one holds suggest will best helpsone achieve one's rationally relevant goal. If this is true, there is no wonder why we say that *it is irrational not to perform one of the actions which the evidence one holds suggest will best help one achieve one's rationally relevant goal*, as this follows by logical necessity. Furthermore, we might think that when we say that *it is irrational for an individual to not

believe what she has all things considered reason to believe* we suppose that the individual has the goal of holding true beliefs, and *believing what one has all things considered reason to believe* tends to be a much better means for achieving this than *not believing what one has all things considered reason to believe* is. One can explain why we are calling both of these things 'irrational', by saying that claims about theoretical rationality are claims which are similar to claims about instrumental rationality.

It is not easy to see what reasons we have to suppose that The Rational Coherence View is correct and, therefore, it is not easy to see that the fact that this view supports The Wide-Scope Interpretation gives us reason to prefer The Wide-Scope Interpretation over The Narrow-Scope Interpretation.

An argument from linguistic plausibility in favor of The Narrow-Scope Interpretation

It seems that The Wide-Scope Interpretation is only linguistically plausible under The Actual Intention Interpretation, and not under The Desire-Based Interpretation, The Value-Based Interpretation, or The Prudence-Based Interpretation. Because the rationally relevant goal of an individual is the best possible fulfillment of her rationally relevant desires provided that she intends this goal under The Desire-Based Interpretation, what constitutes this goal would change only if one's rationally relevant desires were to change given this interpretation. A reason The Wide-Scope interpretation does not seem linguistically plausible under The Desire-Based Interpretation is that it seems natural to suppose that normally when an individual tells another that an action he could perform would be irrational, she means to tell him not to perform the action and does not mean to urge him to change his intrinsic desires. For instance, if I am telling a friend, who has won tickets to meet her idol and has the intrinsic desire to meet the idol, that it would be irrational for her to throw the tickets in the thrash*, it seems that I would mean to advice her *to not throw the tickets in the thrash*, and not mean to advice her *to not throw out the tickets *or* give up on the intrinsic desire to meet her idol*. According to The Wide Scope Interpretation, however, in order to be in accordance with The Instrumental Requirement under The Desire-Based Interpretation, the actor might just as well change her intrinsic desires. If The Instrumental Requirement is to be understood in this way, one would expect that claims about irrationality would be used to urge people to either do something different *or* change their intrinsic desires. Because we commonly do not

seem to mean to do this when we claim that an action would be irrational, this suggests that The Wide-Scope Interpretation is linguistically implausible under The Desire-Based Interpretation.

It seems clear that The Wide-Scope Interpretation would be linguistically implausible under The Value-Based Interpretation and The Prudence-Based Interpretation as well. The reason for this is that the rationally relevant goal according to The Prudence-Based Interpretation, must be what is determined by what is best for the actor on the whole, and according to The Value-Based Interpretation, the rationally relevant goal must be determined by what is independently valuable. These things do not seem like something an individual can change. Therefore, it seems to follow that according to these interpretations, an actor cannot meet The Instrumental Requirement by changing her rationally relevant goal.

In Section 3.1, I pointed out that The Actual Intention Interpretation seems to be linguistically implausible. It, therefore, seems that on the basis that there seems to be no linguistically plausible combination of views which includes The Wide-Scope Interpretation, we have reason to adopt The Narrow-Scope Interpretation, and think that The Instrumental Requirement can only be met by means of performing the actions which the evidence one holds suggest will best help achieve one's rationally relevant goal.

4 - The End Point Pursuit Understanding of rational action

4.0 - Chapter introduction

In the previous chapter, I explained nine different questions which it seems reasonable to ask with regard to how The Instrumental Requirement should be interpreted. For each question I argued for a particular answer to the question. In this chapter, I will formulate an understanding of rational action based on these interpretations, and I will consider whether this understanding seems to hold up in light of several ideas about the concept of rational action which might seem reasonable.

4.1 - Articulating The End Point Pursuit Understanding

In Chapter 2, I suggested the following formulation of The Instrumental Requirement:

The Instrumental Requirement of Practical Rationality

In order for an action X to be rational for an individual to perform in a given situation, it is a requirement that action X is at least as constructive as any other action for achieving the relevant goal of the individual.

In Chapter 2, I also argued that it seems reasonable to pursue a Humean conception of rational action which is the view that an action is rational if and only if it fulfills The Instrumental Requirement. Based on the discussions in Chapter 3, we seem to have reason to suppose that the rationally relevant goal is *the goal of the best possible fulfillment of the totality of the actor's contemporary intrinsic desires, provided that the actor intends to achieve this by performing the given action*. We, therefore, seem to have reason to think that the following understanding, which I will call 'The End Point Pursuit Understanding of rational action for reasons which will be made clear in the next section, seems like a reasonable conception of rational action:

The End Point Pursuit Understanding of rational action

It is rational for a given individual to perform an action X in a given situation if and only if action X is at least as constructive as any other action for fulfilling the totality (given by The Holistic Interpretation) of the relevant desires (given by The Desire-Based Interpretation) of the individual, and she intends to achieve this by performing the action.

1. Where only intrinsic desires are rationally relevant desires, and instrumental desires are not (Given by The Intrinsic Interpretation)
2. Where there are no ways in which an intrinsic desire can be criticized with reference to other intrinsic desires that can make it rationally irrelevant. (Given by The Irrelevance of Criticism Interpretation)
3. Where only contemporary desires are rationally relevant desires, and future desires are not. (Given by The Contemporary Desires Interpretation)
4. Where contemporary desires are understood as including, both occurrent and standing desires. (Given by The Standing Inclusive Interpretation)
5. Where the basis on which an intrinsic desire has been formed does not influence whether or not it is a rationally relevant desire. (Given by The No Laundering Interpretation)
6. Where what best helps fulfill the totality of the rationally relevant desires of an individual is understood as what the evidence the actor holds suggest would best help fulfill the totality of her rationally relevant desires, not what actually would nor what the actor believes would best help fulfill the totality of these desires. (Given by The Evidence-Centered Interpretation)
7. Where The Instrumental Requirement can not be met by changing one's rationally relevant goal to be in accordance with action X. (Given by The Narrow-Scope Interpretation)

This means that The End Point Pursuit Understanding can simply be formulated in this way:

The End Point Pursuit Understanding of rational action

It is rational for a given individual to perform an action X in a given situation if and only if the evidence she holds suggests that it is at least as constructive as any other action for fulfilling the totality of her contemporary intrinsic desires and she intends to achieve this by performing the action.

Given The End Point Pursuit Understanding, an action which is rational means *an action which the evidence of the actor suggests will best help fulfill the totality of her contemporary intrinsic desires performed by the actor with the intention to try to do what will best help fulfill these desires*.

4.2 - Natural End Point

A method for trying to persuade an actor to perform a different action from the one she planned to perform is to point at facts about the world, such as facts about the world outside the actor, facts about what the actor has reason to believe given the information she has, and facts about what desires the actor has. This process will have a natural end point when the actor has perfect knowledge about the world, including about herself such as what desires she has, and including knowledge about what actions will actually help fulfill what desires. At this end, point what the evidence she holds suggests is the action which best helps fulfill the totality of her rationally relevant desires and which action actually would best help fulfill the totality of her rationally relevant desires would be the same. Since she at this end point already has all the facts, pointing at facts would not do anything to persuade the actor. It seems to me that it is reasonable to suppose that, at this end point, actors will tend to be most strongly motivated to perform the action which actually would best help fulfill the totality of their contemporary intrinsic desires. An action that actually is the one which best helps fulfill the totality of the actor's contemporary intrinsic desires I will call 'End Point rational'.

Let me now sum up the reason why I think acting in accordance with what is End Point rational is what an actor would tend to be most strongly motivated to do when having perfect knowledge. Given that The Humean Theory of Motivation is correct, a human can only be motivated to do something on the basis of fulfilling a desire. This makes it reasonable to think that if an individual, having perfect knowledge, was to carefully choose between doing what best helps fulfill her desires, what best leads to what is good for her, what best leads to what is of independent value, and what best helps her achieve what she intended to achieve before having perfect knowledge, she would tend to choose to do what best helps fulfill the totality of her desires. In other words, that she would act in accordance with both The Desire-

Based Interpretation and The Holistic Interpretation. Furthermore, it seems reasonable to think that if an individual having perfect knowledge was given the choice between doing what actually best helps fulfill the totality of her desires and what the evidence she had before having perfect knowledge suggested would best help fulfill the totality of her desires, she would choose the first. In other words, that she would act in accordance with The Objective Interpretation.

Similarly, it seems reasonable to think that if an individual having perfect knowledge was given the choices between doing what best helps fulfill the totality of 1) her intrinsic desires, or both her intrinsic desires and her instrumental desires, 2) her non-criticizable desires, or both her criticizable and non-criticizable desires, 3) her contemporary desires, or both her contemporary desires and her future desires, 4) her occurrent desires, or both her occurrent desires and her standing desires, and 5) her desires that have been laundered in some way, or her non-laundered desires, she would then tend to choose to do what best helps fulfill the totality of 1) her intrinsic desires, 2) both her criticizable and non-criticizable desires, 3) her contemporary desires, 4) both her occurrent and standing desires, and 5) her non-laundered desires, respectively. In other words, that she would choose to act in accordance with The Intrinsic Desires Interpretation, The No Irrelevance of Criticism Interpretation, The Contemporary Desires Interpretation, The Standing Inclusive Interpretation, and The No Laundering Interpretation. Lastly, it also seems reasonable to think that if such an individual was given the choice between doing what best helps fulfill the totality of her rationally relevant desires, or changing her contemporary intrinsic desires to be in accordance with whatever actions she is about to take, she would tend to choose the first. Given that an individual, having perfect knowledge, would tend to act in these ways, she would tend to act in accordance with what is End Point rational.

Now, performing the action which her evidence suggests is the action that best will help fulfill the totality of her contemporary intrinsic desires seems like the best possible strategy available to an individual for maximizing her chances of doing what is End Point rational. Given that a rational action is the action which the evidence of the actor suggest will best help fulfill the totality of her contemporary intrinsic desires, i.e., given that The End Point Pursuit Understanding is correct, it seems reasonable to think that the best possible fulfillment of the totality of her contemporary intrinsic desires is what an actor typically tries

to achieve when trying to do what is rational. In other words, it seems reasonable to think that what typically motivates an actor to try to do what is rational for her to do is the prospect of doing what is End Point rational. This is why I have chosen to refer to the understanding of rational action proposed here as The End Point Pursuit Understanding of rational action.

4.3 - Testing The End Point Pursuit Understanding

If my arguments regarding the linguistic plausibility of the all of the interpretations that make up The End Point Pursuit Understanding are correct, it seems to follow that this view is in accordance with all of the example cases discussed in the previous chapter and that it meets The Fallibility Requirement and the idea that rational action has universal de facto authority. This alone makes The End Point Pursuit Understanding look promising. I will subject the view to some testing by considering a few more ideas one might think that an understanding of rational action must accommodate in order to be convincing and see if it seems that these ideas pose serious challenges to the view.

The idea that a person must care about doing what is rational

In her argument against realist conceptions of The Instrumental Requirement, Korsgaard repeatedly stresses that a theory of instrumental rationality must give an account of "why we must […] care about" the fact that an action is rational, and "explain how that fact gets a grip on the agent" (Korsgaard, 1997, pp.53-54). She asks rhetorically: "suppose I do not care about being rational? What then?" (Korsgaard, 1997, p.56). She argues that "if there is a principle of practical reason which *requires* us to take the means to our ends, then those ends must be […] [ends] that we have some reason to [pursue]" (Korsgaard, 1997, p.64).

It seems that implied in these remarks is the claim that in order to be a plausible understanding of rational action, it must imply that an actor *must* care about the fact that an action is rational. The End Point Pursuit Understanding alone seemingly provides no basis for saying that we *must* care about the fact that an action is rational. If Korsgaard's claim is correct, it seems to be problematic for the view that The End Point Pursuit Understanding is plausible as a complete theory of rational action.

Let us, therefore, consider Korsgaard's claim. It seems clear that people have a strong tendency to act in accordance with what they think is rational. However, that an understanding of rational action must imply, in order to be plausible, that an actor *must* care about the fact that an action is rational, obviously does not follow. If this observation is what the claim is based on this requirement seems unreasonably strict. Given this, it seems that the reasonable requirement is that an understanding of rational action must be compatible with the phenomenon that people have a strong tendency to act in accordance with what they think is rational. Given that this second requirement is the correct one, it is not necessary to explain why one must care about performing actions that are rational.

This second requirement it seems plausible that The End Point Pursuit Understanding is able to meet. It, namely, seems very plausible that 1) when people believe that their evidence suggests that a given action is the one which will best would help fulfill the totality of their contemporary intrinsic desires, then they would have a strong tendency to perform that action, and that 2) the reason people have a strong tendency to do what they believe to be rational is because when they believe that an action is rational, then they believe that the action is the one which their evidence suggests will best fulfill the totality of their contemporary intrinsic desires. This would make The End Point Pursuit Understanding compatible with the observation that people seem to have a strong tendency to act in accordance with what they think is rational.

The idea that a person *ought* to pursue one's rationally relevant goal

In Section 3.8 I of pointed out the psychological distinction between occurrent and standing desires and argued that The Standing Inclusive Interpretation allows for the possibility that people can perform irrational actions because they can fail to do what best helps fulfill the totality of all of their occurrent and standing desires as a whole. Korsgaard acknowledges the possibility to "make [the] psychological distinction between what a person […] locally wants and what he "really wants"." and based on that claim that people can perform actions which are "irrational because they do not promote the ends that they "really want" " (Korsgaard, 1997, p.42). Such a move is similar if not identical to the move made in Section 3.8.

Korsgaard does present an objection to the move she describes (Korsgaard, 1997, p.42). A very similar objection seemingly can be made against my move. I will, therefore, construct

this type of objection against my move on Korsgaard's behalf. We can formulate such an objection by paraphrasing her formulation of the original objection. What is inside square brackets are changes from her original formulation:

> If we are going to appeal to [the fulfillment of all contemporary intrinsic desires] as a basis for making claims about whether people are acting rationally or not, we will have to argue that a person *ought* to pursue [the fulfillment of all her contemporary intrinsic desires] rather than [the fulfillment of her occurrent intrinsic desires]. That is, we will have to accord [the fulfillment of all contemporary intrinsic desires] some normative force. It must be a requirement of reason that you should do what [best helps fulfill the totality of all of your contemporary intrinsic desires], even when you are tempted not to. (Korsgaard, 1997, p.42)

Now, it seems that we would commonly say that one typically ought to do what is rational ,all other things being equal. This suggests that if an understanding of rational action does not imply that one typically *ought* to do what is rational, all other things being equal, in any plausible sense of the word 'ought', then it is linguistically implausible. Let us, therefore, grant that an understanding of rational action must imply that an individual typically ought to do what is rational, according to some plausible conception of the word 'ought'. To see if it is plausible that The End Point Pursuit Understanding can meet this requirement, we must consider whether there is a plausible understanding of the word 'ought' that makes it so that The End Point Pursuit Understanding implies that one typically ought to do what is rational.

It seems plausible to think that we use the word 'ought' to refer to multiple different concepts, similar to the way we sometimes use the word 'letter' to refer to the concept of '"a character of an alphabet"', and other times to the concept of '"a certain type of written message"'. It seems plausible that we use the word 'ought' to refer to different concepts in statements about morality, such as *you ought to do what is morally right*, and statements about rational action such as *you ought to not perform irrational actions*. If this is correct we can distinguish between 'ought' used to refer to what we refer to in the first type of statement, a concept we can call 'morally ought', and 'ought' used to refer to what we refer to in the second type of statement, a concept we can call 'rationally ought', and that while the concepts have some similarities, none of them can be derived from the other. Henceforth, when I write '(rationally) ought' I mean to only refer to the concept we commonly mean to refer to when we use the word 'ought' in statements about rational action.

Given the above, it seems plausible that 'one (rationally) ought' in claims such as *one (rationally) ought to do what is rational* simply means the same as 'it would be rational'. This would mean that *an action one (rationally) ought to perform* simply means *an action it would be rational to perform*. Now, given this meaning of '(rationally) ought' The End Point Pursuit Understanding would, of course, imply that a person (rationally) *ought* to pursue the best possible fulfillment of the totality of her contemporary intrinsic desires. Because of this, and because this understanding of '(rationally) ought' seems plausible, it seems plausible that The End Point Pursuit Understanding meets the requirement that it must imply that one typically ought to do what is rational, all other things being equal.

The idea that one has to provide an argument for why

One might claim that one has to provide an argument for *why* it is rational for an actor to perform the action which The End Point Pursuit Understanding entails that is rational for her to perform, or that 2) one has to provide an argument for *why an actor (rationally) ought* to perform such an action.

Let me start by addressing the first version of this objection. For each of the interpretations I suggested that we should adopt in Chapter 3, I argued that it seemed to be more in line with our use of the term 'rational' to describe actions than any of the other interpretations considered. As a result, The End Point Pursuit Understanding hopefully corresponds rather well to how we use the term 'rational action'. The following response to the above claim, therefore, seems to be available:

It seems like the understanding of rational action as being *action which the evidence of the actor suggest will best help fulfill the totality of her contemporary intrinsic desires, performed by the actor with the intention to do what best helps fulfill these desires* corresponds really well with how we use the term 'rational action'. In other words, this seems to be what we mean to say when we call an action rational. Given that this is true, it seems that it would be misplaced to ask for an argument for why it is rational to perform such an action. That seemingly would be like asking for an argument for why the molecule H_2O is water. The molecule H_2O is water because the molecule H_2O is what we use the term 'water' to refer to. Certainly, an argument for why we use the term 'water' to refer to H_2O is not required to defend the claim that H_2O is water. Similarly, given that the above is true, *an

action which the evidence of the actor suggest will best help fulfill the totality of her contemporary intrinsic desires, performed by the actor with the intention to try to do what best helps fulfill these desires* would be rational because this is what we use the term 'rational action' to refer to.

One can draw on this response to the first version of the objection to address the second version; that one has to provide an argument for *why we (rationally) ought* to perform the action which is rational according to The End Point Pursuit Understanding. Earlier I argued that it is plausible that we commonly use the term 'one (rationally) ought' to mean the same as 'it would be rational'. If it is the case that 1) 'one (rationally) ought' means the same as 'it would be rational', and that 2) what we mean to refer to by 'an action it would be rational to perform' is '"an action which the evidence of the actor suggest will best help fulfill the totality of her contemporary intrinsic desires, performed by the actor with the intention to try to do what best helps fulfill these desires"', then it follows, given that we are consistent in these uses, that I) what we commonly mean to refer to by 'an action one (rationally) ought to perform' is '"an action which the evidence of the actor suggest will best help fulfill the totality of her contemporary intrinsic desires, performed by the actor with the intention to try to do what best helps fulfill these desires"'. Given this, for the same reason it seems misplaced to ask why H_2O is water, it seems misplaced to ask why we (rationally) ought to perform such an action.

Conclusion

In this thesis, I have argued why The End Point Pursuit Understanding of rational action seems reasonable. I started by explaining why it seems reasonable to think that an action must meet The Instrumental Requirement is order to be a rational action, and argued that this requirement can be formulated as the following: In order for an action X to be rational for an individual to perform, given that goal A is the relevant goal, it is a requirement that action X is at least as constructive for achieving goal A compared to all other available actions. I argued that it seems reasonable to pursue The Humean View view that this is the only requirement of practical rationality, as opposed to The Kantian View is the view that there is a second type of requirement, which can be called 'The Categorical Requirement'. The Kantian View, I pointed out, has implications for which actions would be rational to perform that seem implausible, and I argued that, because of this, this view does not seem promising. Next, I identified nine different questions with regard to how The Instrumental Requirement is to be interpreted and argued that we should adopt the interpretations which lead to the understanding of rational action which is the most in line with how we commonly use the terms 'rational', 'rationally required', and 'irrational' to describe actions. I then examined which interpretations seems the most linguistically plausible following this approach. I argued that it seems that we have reason to adopt the view that *an action fulfills The Instrumental Requirement if and only if it is the action which the evidence of the actor suggest will best help fulfill the totality of the contemporary intrinsic desires of the actor, and the actor intends to try to achieve this goal by performing the action*. On this basis, I argued it seems reasonable to think that an action that fulfills this requirement is a rational action and called this view The End Point Pursuit Understanding. I explained that the implications of this understanding seem to correspond with the use of the words related to rational action with regard to all the example cases presented in this thesis and that it seems to be compatible with the idea that one ought to do what is rational all other things being equal, the idea that it is possible to perform irrational actions, and the idea that humans have a tendency to act in accordance with what they believe to be rational. This strengthens the impression that this seems like a promising understanding of rational action.

References

Andreau, C. (2006). “Might Intentions Be the Only Source of Practical Imperatives?,” *Ethical Theory and Moral Practice*, 9(3): 311–325.

Beardman, S. (2007). “The Special Status of Instrumental Reasons,” *Philosophical Studies*, 124(2): 255–287.

Bratman, M. (2009). “Intention, Belief, and Instrumental Rationality,” in D. Sobel and S. Wall (eds.), *Reasons for Action*, Cambridge: Cambridge University Press, pp. 13–36.

Broome, J. (1997). “Reason and Motivation,” *Proceedings of the Aristotelian Society* (Supplementary Volume), 71: 131–146.

Broome, J. (1999). ‘Normative Requirements’, *Ratio*, 12: 398–419.

Gauthier, D. (1986). *Morals by Agreement*, Oxford: Clarendon Press.

Jackson, F. & Pargetter R. (1986). “Ought, Options, and Actualism,” *Philosophical Review*, 95(2): 233–255.

Jollimore, T. (2005). “Why Is Instrumental Rationality Rational?,” *Canadian Journal of Philosophy*, 35(2): 289–308.

Kant, I. (1785). Groundwork of the Metaphysics of Morals [G]. In Kant’s Practical Philosophy, trans. Mary J. Gregor. Cambridge: Cambridge University Press, 1996

Korsgaard, C. (1997). “The Normativity of Instrumental Reason,” in G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*, Oxford: Oxford University Press, pp. 215–254

Kelly, T. (2003). “Epistemic Rationality as Instrumental Rationality: A Critique,” *Philosophy and Phenomenological Research*, 66(3): 612–640.

Kolnai, A. (2001). ‘Deliberation is of Ends’, in *Varieties of Practical Reasoning*, E. Millgram (ed.), Cambridge, Mass.: MIT Press.

Kolodny, N. (2005). ‘Why be Rational?’, *Mind* 114: 509–63.

Kolodny, N. (2007). “How does Coherence Matter?,” *Proceedings of the Aristotelian Society*, 107(1–Part 3): 229–263.

Kolodny, N. (2008). “Why be Disposed to Be Coherent?” *Ethics*, 188(3): 437–463.

Kolodny, Niko & Brunero, John. (2018). "Instrumental Rationality” [PDF file], In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018 ed.). Retrieved from https://plato.stanford.edu/archives/win2018/entries/rationality-instrumental/

Lavin, D. (2004). ‘Practical Reason and the Possibility of Error’, *Ethics*, 114: 424–57.

Lord, E. (2014). “The Coherent and the Rational,” *Analytic Philosophy*, 55(2): 151–175.

Lord, E. (2017). “What You’re Rationally Required to Do and What you Ought to Do (Are The Same Thing!),” *Mind*, 126(504): 1109–1154.

Millgram, E. (1995). ‘Was Hume a Humean?’, *Hume Studies*, 21: 75–93.

O'Neill, O. (1989). ‘Consistency in Action’, in her *Constructions of Reason*, Cambridge: Cambridge University Press.

O'Neill, O. (2004). Chapter 6 KANT: Rationality as Practical Reason. In A. R. Mele, & P. Rawling (Eds.), *The Oxford Handbook of Rationality* (pp.93-109). Oxford, United Kingdom: Oxford University Press. (Original work published 2004)

Rippon, S. (2011). "In Defense of the Wide-Scope Instrumental Principle," *Journal of Ethics and Social Philosophy*, 5(2): 1–21.

Rosati, Connie S. (2016). "Moral Motivation" [PDF file]. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). Retrieved from https://plato.stanford.edu/archives/win2016/entries/moral-motivation/

Schroeder, Tim. (2015). "Desire" [PDF file]. In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 ed.). Retrieved from https://plato.stanford.edu/archives/sum2017/entries/desire/

Shpall, S. (2013). "Wide and Narrow Scope," *Philosophical Studies*, 163: 717–736.

Smith, M. (1987). "The Humean Theory of Motivation," *Mind*, 96: 36–61.

Smith, M. (2004). Chapter 5 Humean Rationality. In A. R. Mele, & P. Rawling (Eds.), *The Oxford Handbook of Rationality* (pp.75-92). Oxford, United Kingdom: Oxford University Press.

Smith, M. (2004). "Instrumental Desires, Instrumental Rationality," *Proceedings of the Aristotelian Society* (Supplementary Volume), 78(1): 93–109.

Valaris, M. (2014). "Instrumental Rationality," *European Journal of Philosophy*, 22(3): 443–462.

Wallace, R. Jay. (2014). "Practical Reason" [PDF file]. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018 ed.). Retrieved from https://plato.stanford.edu/archives/spr2018/entries/practical-reason/

Way, J. (2011). "The Symmetry of Rational Requirements," *Philosophical Studies*, 155(2): 227–239