

**The London School of Economics and Political
Science**

Reasons, Rationality and Preferences

Stuart Yasgur

A thesis submitted to the Department of
Philosophy, Logic and Scientific Method of the
London School of Economics for the degree of
Doctor of Philosophy, London, September 2011

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of <55,144> words.

Abstract

The theory of choice receives formal treatment in decision theory, game theory and substantial parts of economics. However there is cause for concern that the formal treatment of the subject has advanced beyond the substantive grounds on which it relies. For, the formal theories fundamentally rely on a concept of preference, which is itself lacking a viable substantive interpretation.

Indeed the challenges to the substantive interpretation of 'preference' threaten to undermine the standard arguments used to justify the completeness and transitivity conditions on which Preference Theories rely.

This discussion will explore whether a conception of rationality, anchored in a larger conception of practical reasoning, can justify the completeness and transitivity conditions.

Specifically, this dissertation will draw on recent developments in philosophy of law, action theory and ethics to enumerate a conception of practical reasoning that takes reasons to be the basic normative concept. It will then seek to offer an account of rationality that is distinct from, but complementary to, the role of reasons. And from this foundation develop an account of preferences that includes many of the characteristics of standard accounts, yet is situated within this broader context. From this vantage point, the discussion will explore possible justifications for the completeness and transitivity conditions. Ultimately, it will be argued that both can be justified – though with different force – in specified domains.

While the discussion will primarily focus on the justification of the completeness and transitivity conditions, it is in part motivated by the goal of exploring the connections between the treatment of choice in the distinct fields associated with Preference Theories and action theory broadly defined. In so doing, the hope is to suggest that there is promise in drawing together formal and substantive treatments of choice which is deserving of greater attention.

Table of Contents

Acknowledgements	5
Chapter 1: Introduction	6
Chapter 2: Reasons	51
Chapter 3: Rationality	129
Chapter 4: Preferences	192
Chapter 5: Conclusion	223
Bibliography	227

Acknowledgements

To Sirine, words do not begin to acknowledge my gratitude. Let me simply say, without whom not.

To my partners in crime, Sheldon and Mauro, thank you for challenges, the new ideas, and most of all for making this a pleasure.

To my parents, for tireless support and taking this on as your own.

To ECK and DCK, for the endless inspiration of living life well.

To my new family, thank you for that rarest of gifts - a supportive sanctuary.

To Richard, whose grace and generosity is an example I will learn from far beyond philosophy.

Chapter 1: Introduction

The theory of choice receives formal treatment in decision theory, game theory and substantial parts of economics. However there is cause for concern that the formal treatment of the subject has advanced beyond the substantive grounds on which it relies. For, the formal theories fundamentally rely on a concept of preference, which is itself lacking a viable substantive interpretation. This poses a dual edge challenge to Preference Theories, as they will be called here. For on the one hand without a viable substantive interpretation it is difficult to understand the empirical significance of these prodigious formal theories, and on the other hand there is little basis to understand how the theories themselves should change to account for the realities of their subject matter.

This is the challenge that motivates this dissertation. The source of this challenge can be stated concisely.

Decision theory, game theory and substantial parts of economics regularly assume that rational preferences are complete and transitive.

The difficulty is that there have long been questions about how to interpret the concept 'preference'. And, whether the completeness and transitivity condition are justified given these interpretations.

Indeed, Mandler argues that the justification of the completeness condition relies on understanding preferences as choices, while the justification of the

transitivity condition relies on understanding preferences as welfare judgments. Poignantly he argues that preferences can either justify the completeness condition or the transitivity condition, but not both.

For those theories that rely on preferences to be complete and transitive, this is a challenge that cannot be ignored.

Rather than take issue with Mandler's arguments¹ this dissertation will explore a different alternative. At least three possibilities suggest themselves and/or have been pursued before:

1. Alter the formalism to make do with completeness or transitivity, but not both.²
2. Acknowledge the lack of justification for the transitivity and completeness condition, yet stipulate both conditions as idealizations.³
3. Seek a justification for the completeness and transitivity conditions based on alternative substantive grounds.⁴

¹ If anything, Mandler's argument understates the challenge. The discussion will return to this point shortly.

² While this is a viable alternative, it will not be pursued here for two reasons. First, the success of Preference Theories suggests there may be reason to preserve the formalism intact, if possible. And second, even if it is possible to address the current challenge by jettisoning either the completeness or transitivity conditions, it leaves open the question of how to interpret 'preference'.

³ It is undoubtedly true that there is a degree of idealization in Preference Theories. Yet relying on this fact too broadly can limit Preference Theories' ability to address interesting questions such as: how do, and should, agents choose in a normatively complex world characterized by uncertainty?

There is at least one approach which maintains that Preference Theories are idealizations yet does not walk away from these questions. On this approach Preference Theories describe the choices of idealized agents, yet it is acknowledged that actual agents are somewhat more limited and therefore may systematically deviate from the ideal.

While this type of approach has garnered significant interest, and may prove useful, it is not the approach that will be followed here. For it seems *prima facie* troubling for an account of choice to be based on a view that agents are limited because of their failure to live up to an unjustified ideal.

⁴ This is one way to understand Broome's focus on the betterness relation. The discussion will return to this point shortly.

This current effort will explore the third of these alternatives. Specifically, this discussion will explore whether a conception of rationality, anchored in a larger conception of practical reasoning, can justify the completeness and transitivity conditions.

This may initially seem a cause for concern, because Preference Theories are often lauded for the limited grounds on which they rely. However, parsimony is only one of the theoretical virtues that such theories may embody. And the evidence suggests that Preference Theories do not embody the virtue of parsimony to the extent previously thought. This raises questions of what other theoretical values may be realized.

Whether this account is ultimately successful will not only depend on its ability to offer consistent justification of the completeness and transitivity conditions, but also on the extent to which it realizes other theoretical virtues.

The following discussion will be preliminary in many respects. It will seek to draw from current literature to provide an articulation of an emerging conception of practical reasoning, but will only discuss basic aspects of that view. Further, it will offer possible justifications of the completeness and transitivity condition, but will inevitably leave many questions of theoretical fruitfulness to later exploration.

Nonetheless, the aspiration is that this discussion will point towards a worthwhile avenue for further exploration.

--

This chapter is organized in two stages. The first stage discusses the challenges facing Preference Theories; and the second describes the approach to those challenges that will be pursued in subsequent chapters.

The first stage is comprised of three sections. The first section succinctly describes Revealed Preference Theory and Sen's critique of it to illustrate the long legacy of tension between interpretations of 'preference' and the justification of consistency conditions which apply to them. The second section turns to Mandler's argument which establishes that rationality cannot justify both the completeness and transitivity conditions. And, the third section builds on this point by suggesting that Mandler may have understated the challenges to justifying the completeness and transitivity conditions in three ways:

- the completeness condition is not a condition of rationality;
- the transitivity condition is not justified by the standard argument used to do so;
- the transitivity condition faces the threat of being vacuous.

Recognizing the force of these challenges, the second stage introduces the approach that will be employed for the remainder of this dissertation for exploring a positive account of preference that can offer a (qualified) justification of the completeness and transitivity conditions. The second stage is also comprised of three sections. The first briefly describes three observations that will serve as a point of departure for the subsequent discussion. The second

offers reflections on the nature of the current inquiry to suggest a broader scope for the discussion. And the third, briefly lays out the strategy which will be pursued.

Stage 1: The Challenges

Decision Theory, Game Theory and economics regularly assume that rational preferences are complete and transitive.⁵

Simplistically put, the transitivity condition reflects the idea that if an agent prefers an alternative, x , to another alternative, y , and he also prefers the second alternative, y , to a third alternative, z , then the agent also prefers the first alternative, x , to the third alternative, z . Further, once the preference relation has been defined, transitivity can be easily formalized. Take the preference relation to be denoted by the symbol \succeq . Following Mandler, the expression $x \succeq y$ means that the agent prefers x to y or is indifferent between the two. Strict preference and indifference are defined in terms of \succeq : x is strictly preferred to y , denoted $x \succ y$, if $x \succeq y$ and it is not the case that $y \succeq x$, and x and y are indifferent if both $x \succeq y$ and $y \succeq x$.⁶ A preference relation \succeq is defined to be transitive if, for all triples of consumption bundles (x, y, z) , $x \succeq y$ and $y \succeq z$ imply $x \succeq z$.⁷

⁵ Since this is well established, it will not be discussed in detail here. For a detailed discussion of completeness and transitivity in cardinal utility theories see Von Neuman and Morgenstern. For the ordinal representation theorem that proves that complete and transitive preferences can, with some additional conditions, be represented by a continuous utility function see Debreu 1959, pg 55 – 9.

⁶ Mandler “A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both”, pg 9.

⁷ For consistency throughout this chapter, Mandler’s formal definition of transitivity in “A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both” has been used. Other definitions could have been used as well.

Similarly, a simple way to understand completeness is that the completeness condition reflects the idea that an agent can rank any pair of alternatives.

Further Completeness can be formalized as: a preference relation \succeq is defined to be complete if, for all pairs of consumption bundles (x, y) , either $x \succeq y$ or $y \succeq x$ (or both).⁸

The difficulty is that there have long been questions about how to interpret the concept 'preference', such that the relevant consistency conditions are justified. This can be illustrated through the example of Revealed Preference Theory, an early articulation of Preference Theories.

Revealed Preference

In Revealed Preference Theory which was introduced by Samuelson⁹, 'preference' was interpreted as choices. In part the motivation for Revealed Preference Theory was to explain an agent's behavior through observations of her other behaviors; and in so doing to avoid the difficulty of attributing troublesome mental states such as desires and beliefs, to the agent.¹⁰ The basic idea is that: "If a collection of goods y could have been bought by a certain individual within his budget when he in fact was observed to buy another collection x , it is to be presumed that he has revealed a preference for x over y .

⁸ Mandler, M. "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 10.

⁹ P. A. Samuelson (1938) "A Note on the Pure Theory of Consumer's Behaviour", *Economica*, v. 5: pp. 61-71. P. A. Samuelson (1948) "Consumption Theory in Terms of Revealed Preference", *Economica*, v. 15: pp. 243-253. P. A. Samuelson (1950) "The Problem of Integrability in Utility Theory", *Economica*, v. 17: pp. 355-385.

¹⁰ In "A Revision of Demand Theory" by J.R. Hicks, (Oxford, 1956), Page 6, the following illustrative quote citing this motivation and the understood significance of revealed preference theory is offered: "the econometric theory of demand does study human beings, but only as entities having certain patterns of market behavior; it makes no claim, no pretence, to be able to see inside their heads."

The outside observer notices that this person chose x when y was available and infers that he *preferred* x to y .¹¹

To extend this approach to more than a single choice Samuelson developed the Weak Axiom of Revealed Preference (WARP) as a consistency condition¹², which stipulates roughly that if an individual shows a preference for x over y in the way just described, then she cannot also reveal a preference of y over x .¹³

Unfortunately, as Sen argued in “Behavior and the Concept of Preference”¹⁴, there is a tension in Revealed Preference from the outset. For the justification of WARP depends on attributing preferences to the agent under consideration, which is one of the complications Revealed Preference was supposed to avoid. The reason for this is straightforward. If preferences, understood as attitudes or judgments that underlie choices, are not attributed to the agent under consideration, then WARP would simply require that the agent’s choices be consistent. Without appealing to something other than choices there is no justification for requiring that an agent’s choices be consistent. However, if preferences are attributed to an agent, and the agent’s preferences remain constant, then if the agent’s choices are to be consistent with the agent’s preferences, the agent’s choices must also be consistent. While attributing preferences to the agent provides adequate grounds to justify the consistency condition, failing to do so undermines the justification for WARP on which Revealed Preference depends.

¹¹ , A.K. Sen, *Choice Welfare and Measurement*, (First Harvard University Press, 1998) Pages 54-55

¹² P. A. Samuelson, “A Note on the Pure Theory of Consumer’s Behavior”, *Economica*, 5 (1938)

¹³ For the sake of readability this discussion relies on the version of WARP put forth by Sen in *Choice Welfare and Measurement* on pg 55.

¹⁴ A.K. Sen, *Choice Welfare and Measurement*, (First Harvard University Press, 1998), Pages 54-73

So even in this early articulation of Preference Theory there was a question of how to interpret 'preference', and this question posed a challenge for the justification of the key consistency condition on which it depends.

It should be noted that questions of how to interpret 'preference' were not restricted to Revealed Preference Theory. Sen continued to draw attention to the prevalence of different interpretations of 'preference' in the literature. For example, "Certainly, there is no remarkable difficulty in simply defining preference as the underlying relation in terms of which individual choices can be explained . . . In this mathematical operation preference will simply be the binary representation of individual choice. The difficulty arises in interpreting preference thus defined as preference in the usual sense with the property that if a person prefers x to y then he must regard himself to be better off with x than with y."¹⁵ And, "the normal use of the word permits the identification of preference with the concept of being better off, and at the same time it is not quite unnatural to define "preferred" as "chosen". I have no strong views on the "correct" use of the word "preference", and I would be satisfied as long as both uses are not simultaneously made, attempting an empirical assertion by virtue of two definitions."¹⁶¹⁷

Justification for the completeness and/or transitivity condition

¹⁵ Sen, *On Economic Inequality* 1973.

¹⁶ Sen, *Rational Fools* 1977.

¹⁷ The two interpretations of 'preference' that Sen points to have been noted by economists and philosophers. They have been referred to by a variety of labels and slightly different definitions. For example, in "Sympathy, Commitment and Preferences", Hausman refers to these understandings of preference as 'expected advantage ranking' and 'choice ranking' respectively. (Pg 34) And in "A difficult choice in preference theory: rationality implies completeness or transitivity but not both", Mandler refers to the welfare definition and choice definition of preferences respectively. (Pg 15)

In “A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both”, Mandler makes a significant contribution by showing that the different interpretations of ‘preference’ do not merely cause difficulties for Revealed Preference Theory, but for Preference Theories more generally. For, as the title of his article suggests, depending on the interpretation of ‘preference’ relied on, rationality can either justify the completeness condition or the transitivity condition, but not both.

Mandler’s argument consists of two different claims that build on each other. The first claim is that there are standard justifications for the completeness condition and the transitivity condition which rely on different interpretations of ‘preference’. The second claim is that the justifications for the completeness and transitivity conditions are incompatible with one another, in the sense that they cannot both be based on a common interpretation of ‘preference’. These two claims will be looked at in turn.

The standard arguments for the completeness condition, which Mandler considers the strongest available, are based on the notion of preference as choice. The basic idea is that any agent whose preferences are possibly incomplete between any two alternatives, x and y , can be forced to choose between x and y by putting the agent in a choice situation such that if he/she does not choose between x and y , he/she will end up with an alternative, z , which is much less preferred to either x or y . Since z can be made sufficiently unattractive, choices will always result. By identifying these choices with preferences, it is claimed that preference orderings are never incomplete.

It is important to note that the choice of x or y that occurs because of this 'forcing procedure' does not reflect a welfare judgment of x over y or y over x .¹⁸ In contrast, the most prevalent justification for the transitivity condition, which comes from the Money Pump argument, relies on the interpretation of preferences as welfare judgments.

Mandler describes the Money Pump argument in the following way: "Here agents exhibit a more blatant violation of transitivity: for some triple of options (x, y, z) preferences satisfy $x > y$, $y > z$, and $z > x$. Because each of these preferences is strict, such an agent, when originally endowed with z , will agree to part with a small amount of money to switch to y , then pay more money to switch from y to x , and then pay more money still to return to z , thereby ending up with the original status quo but with less money. If the judgments $x > y$, $y > z$, and $z > x$ are not altered by the loss of wealth, the agent can be subjected to more rounds of pumping."¹⁹

In order to establish a justification for the transitivity condition as a condition of rationality, an understanding of the relevant conception of rationality is required.

Accordingly, Mandler offers a general statement of the relevant conception of rationality. "The claims of preference theory are also less ambitious than is sometimes supposed. Economic analysis does not assert the absurdity that agents always choose the preference maximizing action. The theory claims only

¹⁸ It is interesting to question the force of the resulting claim. For example, does Mandler's conclusion offer a justification of the completeness condition as a condition of rationality or merely a positive claim that agents' preferences will always be complete? This question will be raised again later.

¹⁹ Mandler, "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 17.

that when agents systematically violate the dictates of economic rationality – which posit that agents can rank any pair of options and that rankings are transitively ordered – they suffer harm. Consequently, given practice and opportunity to learn, their behavior will in time conform more closely to the axioms of rationality. For many, this long-run link to behavior explains the role of rationality in preference theory: rationality can be a force that ultimately guides action.”²⁰

This general statement can be sharpened. For Mandler there is an intimate connection between rationality and harm, such that if an agent systematically violates the dictates of rationality, the agent will suffer harm. Therefore, if one can show that agents who have preferences with certain features suffer harm, then one can justify considering those features of the agents’ preferences irrational.

This can be codified into what may be called a Consequentialist Justification comprised of two elements:

- Consequentialist Conditional: If agents, who have preference orderings with certain features come to harm, then those features are irrational.
- Consequentialist Claim: Agents whose preference orderings have certain features, such as intransitivity, suffer harm.

The Money Pump argument satisfies the Consequentialist Justification by satisfying both the Consequentialist Claim and the Consequentialist Conditional. I.e., The agent ends up with the “original status quo but with less

²⁰ Mandler, “A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both”, pg 1-2

money”²¹. And he does so as a result of the intransitivity of his preferences.

According to the Consequentialist Justification, this provides grounds for considering intransitive preferences irrational. I.e., the transitivity condition is a condition of rationality.

Strictly speaking, Mandler’s argument falls short of establishing that the justification of the transitivity condition relies on an interpretation of ‘preferences’ as welfare judgments. While the claim that the agent suffered harm is a judgment about the agent’s welfare, it is not dependent on interpreting the agent’s preferences as welfare judgments. This is easily remedied.

Mandler’s discussion of the transitivity condition is concerned with the ‘Ordinal Theory of Preference’, in which there are no external standard regarding what is good or bad for the agent. The agent is authoritative over his own ‘good’. To establish that the agent has been harmed one turns to the agent’s own judgments about his welfare. In order to establish that the agent in the Money Pump argument has been harmed requires interpreting his preferences as welfare judgments and showing that the agent ends up in a situation he prefers less to the original, as would be the case if the agent had a preference for more money rather than less.

Therefore, the Money Pump argument does offer a justification of the transitivity condition as a condition of rationality based on the Consequentialist Justification and interpreting the agent’s preferences as welfare judgments.

²¹ Mandler, “A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both”, pg 17.

Taken together, this constitutes Mandler's first claim, that there are standard justifications for the completeness condition and the transitivity condition which rely on different interpretations of 'preference'. The completeness condition relies on interpreting 'preferences' as choices, and the transitivity condition relies on interpreting 'preferences' as welfare judgments.

With this review of the standard justification of the completeness and transitivity condition, the discussion can now turn to Mandler's second claim, that the justifications for the completeness and transitivity conditions are incompatible, in the sense that they cannot both be based on a common interpretation of 'preference'.

There are two alternatives, either the completeness and transitivity conditions can both be justified based on interpreting 'preference' as choice or as welfare judgments.

To begin, consider the possibility of justifying the completeness and transitivity conditions based on interpreting 'preference' as welfare judgments. Since the Money Pump Argument that justified the transitivity condition relied on interpreting 'preference' in this way, the question is whether this interpretation can also offer a justification of the completeness condition.

Unfortunately, the 'forcing procedure' does not offer a justification of the completeness condition if 'preference' is interpreted as welfare judgments. To see this, consider an agent that is presented with the 'forcing procedure' described above, and declines to choose x or y and so ends up with z. Such an

agent would end up with an alternative that he preferred less to either of the other alternatives that were also available.

If the agent's preferences are interpreted as welfare judgments, then the agent would have suffered harm. The Consequentialist Claim is satisfied. However, the feature of the agent's preferences that this example points to is the agent's choosing contrary to his preferences. I.e., contrary to his preference for x over z or his preference for y over z, if the agent can be said to have chosen at all, he chose z over either x or y. The Consequentialist Justification offers a justification of choosing in line with one's preferences. But even in the case in which the agent chose in line with his preferences, by either choosing x or y, to attribute to the agent a preference, interpreted as a welfare judgment, between x and y, would be to over-reach. Interpreting 'preference' as welfare judgments does not offer a justification of the completeness condition based on the standard argument.

That leaves the question of whether interpreting 'preference' as choice would fair better. Since the 'forcing procedure' justifies the completeness condition based on interpreting 'preference' as choice, the question is whether this interpretation can offer a justification of the transitivity condition.

To explore this possibility, Mandler offers the following variation of the earlier example. "When preference is defined as choice we may interpret the expression $a \succeq b$ to mean "out of the set $\{a, b\}$, a is chosen" and $a \succ b$ to mean "out of the set $\{a, b\}$, a is chosen and b is not." If we assume that at least one element is chosen out of every set – in accordance with the forcing procedure –

then this preference-as-choice relation must be complete. Consequently, a violation of transitivity implies there is a triple (x, y, z) that satisfies $x \geq y$, $y \geq z$, and $z > x$. We now deploy the same sequence of exchanges used earlier: if z is the original status quo, the agent will agree to switch to y and then to x .²²

The agent's preferences are intransitive. And through a series of exchanges, the agent can "end up with an option that is never chosen directly over the original status quo."²³ However, as it is this example does not satisfy the Consequentialist Claim. Since no welfare judgments are ascribed, there is no basis for asserting that the agent with intransitive preferences suffers harm.

To address this, Mandler suggests introducing "sliver of psychological content will bridge the gap. If we suppose that $a > b$ implies that the agent judges himself or herself to be better off with a than with b , then we may conclude that intransitive choosers are irrational: they end up with x even though they judge z to be superior."²⁴

This revised interpretation of the strict preference relation provides the basis for satisfying the Consequentialist Claim. For the agent ends up with an alternative he regards to be inferior to the status quo, and so can – on one understanding – be said to suffer harm.

²² Mandler, "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 17.

²³ Mandler, "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 17.

²⁴ Mandler, "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 17.

However, as Mandler points out, this argument relies on the strong reading of the weak preference relation such that “ $a \succeq b$ means that an agent will always choose a from the set $\{a, b\}$. And if the agent is unable to arrive at welfare judgments between these alternatives, this reading is too artificially demanding. An agent in this situation may well display status quo bias, i.e., choosing to stay with an alternative until offered a different alternative which the agent judges to make him better off. In such cases, an agent’s intransitive preferences may not lead the agent to suffer harm.

Incorporating these points in the current example, consider the case in which “ y is unranked in welfare terms relative to both x and z , but... z is ranked superior to x .”²⁵ If z is the status quo again, the agent will not switch to y , and as a result, will not suffer harm. The agent has intransitive preferences, yet does not suffer harm. The Consequentialist Claim is not satisfied, and therefore the Consequentialist Justification is not satisfied. This interpretation of preferences as choices with a “sliver of psychological content” does not offer a justification of the transitivity condition.

Thus, Mandler concludes there is a tension between the justification of the transitivity and completeness conditions.²⁶ “Taking a bird’s eye view of the various arguments in favor of the ordinalist theory of rationality, a curious symmetry in their flaws appears. If preference is defined as a set of welfare judgments, then rational agents will satisfy transitivity but need not obey completeness; if preference is defined as choice, then although agents will

²⁵ Mandler, “A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both”, pg 17.

²⁶ Importantly for present purposes, Mandler also argues that these results reflect challenges with applying Preference Theories across varied domains. This is a suggestion that will come up again later.

definitionally satisfy completeness, rationality does not imply that they must obey transitivity.”²⁷

Mandler poignantly focuses attention on a challenge to Preference Theories. The fact that ‘preference’ is open to multiple interpretations raises questions about the justification of the completeness and transitivity conditions.

Framing the challenge

Mandler’s title concisely frames the issue: rationality justifies completeness or transitivity, but not both. The main thrust of Mandler’s argument is persuasive. He has successfully shown that there is a need to revisit the justification of the completeness and transitivity conditions. And further that successful justifications of both conditions should rely on consistent interpretations of ‘preference’.

While the subsequent discussion is sympathetic to Mandler’s argument, this can be obscured by points of difference. For example, contrary to Mandler, it will be argued that the standard arguments do not justify the completeness and transitivity conditions as conditions of rationality. And a suggested approach for justifying both conditions based on a consistent interpretation of ‘preference’ will be offered. This seems to directly conflict with Mandler’s first and second claim. While it is important to note the difference, too much should not be made of it. For the current discussion takes on board Mandler’s main contribution in accepting that the standard arguments are based on incompatible interpretations of ‘preference’ and do not jointly justify the completeness and

²⁷ Mandler, “A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both”, pg 20.

transitivity conditions. If anything, the current account may be a sharpening of Mandler's work.

There are three main points of difference with Mandler's account. The first point is really a clarification. While Mandler initially suggests that the completeness condition is justified by rationality, the 'forcing procedure' he turns to in order to explain this does not offer such a justification. Mandler seems to acknowledge this in characterizing the situation in the following way: "agents will definitionally satisfy completeness".²⁸ It will be suggested that this is a feature of the completeness condition itself and not merely the result of the 'forcing procedure' used to justify it. Indeed, as a result, it may be fruitful to consider the completeness condition an idealization rather than a condition of rationality.

The second point directly conflicts with arguments Mandler offers. The Money Pump is widely regarded as the justification for the transitivity condition, and Mandler argues for this view. However, it will be argued that the force of the Money Pump is often overstated, and that there is reason to question whether it offers any justification for the transitivity condition. This leaves the justification of the transitivity condition an open question which will be addressed in subsequent chapters.

The third point of difference introduces a challenge to offering a justification of a meaningful transitivity condition that Mandler did not consider. Broome argued persuasively that in the absence of additional rational requirements the

²⁸ Mandler, "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 20.

prospect of fine individuation threatens to make the transitivity condition vacuous. While this is a distinct issue from those raised by Mandler, it goes to the rational justification of the transitivity condition and will be dealt with alongside them.

Taken together, this suggests the challenge can fruitfully be re-framed as follows: can a consistent interpretation of 'preference' offer a justification of the completeness condition and a meaningful transitivity condition?

This section will expand on the challenges of offering an answer to this challenge by exploring each of the three points of difference with Mandler's account.

The completeness condition as a condition of rationality?

The first issue is relatively straightforward. While Mandler initially described the completeness condition as a condition of rationality, his argument does not support this claim. As was already seen, Mandler relies on the Consequentialist Justification to justify conditions of rationality. The Consequentialist Justification in turn relies on an interpretation of 'preference' as welfare judgments in order to satisfy the Consequentialist Claim. The 'forcing procedure', on the other hand, relies on an interpretation of 'preference' as choice. Therefore, the 'forcing procedure' does not satisfy the Consequentialist Claim, and as a result, the Consequentialist Justification does not justify considering the completeness condition a condition of rationality.²⁹

²⁹ Even if a 'sliver of psychological content' is imputed to 'preference' the Consequentialist Justification would justify considering choosing contrary to one's preferences irrational, but it would not justify considering incomplete preferences irrational.

It is possible to look for other ways to justify the completeness condition as a condition of rationality, but this is likely the wrong place to look. For, on the face of it, it would be surprising to find that the completeness condition is a condition of rationality. The reason for this is simple. Taken literally, the completeness condition is extremely demanding. There are an infinite number of alternatives over which an agent may form a preference, and whether 'preference' is interpreted as choice or welfare judgments, it seems eminently plausible that an agent may be unaware of a possibility and therefore may rationally fail to form a preference between two or more alternatives.

Moreover, there is little pretense that the completeness condition is actually a condition of rationality. As Mandler says the 'forcing procedure' shows that "agents will definitionally satisfy completeness"³⁰ Mandler is not suggesting that agents' actual preferences are complete, but that agents can be 'forced' to make choices between any two alternatives. And if 'preference' is interpreted as choice, then agents can be 'forced' to have preferences between any two alternatives. Given this, it may not be objectionable to stipulate that agents' preferences are complete. This claim is an idealization for the sake of theoretical tractability, not a claim about the requirements of rationality.

This raises an interesting question. If the completeness condition is an idealization rather than a rationality condition, what does this do to the significance of Mandler's argument?

³⁰ Mandler, "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 20.

Mandler's argument was based on the observation that the justification of the completeness condition and the transitivity condition relied on different interpretations of 'preference'. If the completeness condition is understood as an idealization rather than a condition of rationality, it should not be surprising that the two conditions have very different justifications.

Nonetheless, Mandler's argument still has force. For it is reasonable to expect that even if the completeness condition is an idealization, it should apply to the same interpretation of 'preference' as the transitivity condition. For example, if the transitivity condition is interpreted as welfare judgments and if those preferences are to be complete and transitive, then it would seem that there must be a justification for claiming that the completeness condition applies to 'preference' interpreted as welfare judgments, even if as an idealization. The 'forcing procedure' does not provide such a justification. Is there another justification for the completeness condition that is based on the same interpretation of 'preference' that justifies the transitivity condition?

This question will be explored in more detail in the positive section of this dissertation in Chapter 4: Preferences. For now it is important to note that the justification of the completeness condition is an open question.

Does the Money Pump Argument justify the transitivity conditions?

The Money Pump Argument is widely regarded as justifying the transitivity condition, but there are reasons to question this view.

This may be surprising given the discussion of the Consequentialist Justification above. As was mentioned, the Money Pump Argument satisfies the Consequentialist Claim, and therefore if the Consequentialist Conditional holds, the Money Pump Argument offers a justification of the transitivity condition. The issue is that there is reason to question the Consequentialist Conditional.

Consider a generalized version of the Consequentialist Justification:

- P1: If X leads an agent to suffer negative consequences, then X is irrational.
- P2: X leads an agent to suffer negative consequences, in suitable circumstances.
- C: Therefore X is irrational.

The Money Pump Argument also satisfies this general version of the Consequentialist Justification. I.e., the Money Pump offers a case in which P2 holds, and if P1 holds, then the conclusion follows. Next, notice that if we take X to be 'false beliefs', then P2 would also hold. If P1 holds, then it would follow that having false beliefs is irrational. Having false beliefs is not irrational, therefore P1 does not hold; and therefore the Money Pump does not offer a Consequentialist Justification for considering intransitive preferences irrational.

Further, because of the gap between preferences and the consequences of choices based on them, there does not seem to be a way to refine P1 so that it would apply to intransitive preferences but not false beliefs.³¹ Of course it is

³¹ Consider the following example:

- P1': If X leads an agent to suffer negative consequences even when he is ideal in every other way, then X is irrational.
- P2': X leads an agent to suffer negative consequences in suitable circumstances, even when he is ideal in every other way.
- C: Therefore X is irrational.

possible to restrict P1 so it only applies to preferences, but such a change would be a mere ad-hoc addition with little grounds other than preserving the Consequentialist Justification of the transitivity condition.

If the Money Pump Argument fails to justify the transitivity condition based on the Consequentialist Justification, this is because of challenges with the Consequentialist Justification, not the Money Pump Argument.³² Perhaps the Money Pump Argument can justify the transitivity condition in another way.

The Money Pump example first appeared in a paper by Davidson, McKinsey and Suppes "Outlines of a Formal Theory of Value, I"³³, as an illustration of the justification they offered for the transitivity condition. Unfortunately, this is of little help for current purposes. The justification of the transitivity condition offered by Davidson, McKinsey and Suppes would not rule out intransitive preference in many cases in which the transitivity condition is thought to apply. And, second, the Money Pump example is not an illustration of the justification offered in the paper. Both these points can be seen quite quickly.

Davidson, McKinsey and Suppes were concerned with a choice over a triple of alternatives, x, y, and z. They argue that if an agent had intransitive

Since P2` still holds for false beliefs, P1` should as well. But it does not.

³² There is a variation of the Consequentialist Justification which arguable would apply to preferences, but not false beliefs. Unfortunately, it fails for other reasons.

Briefly put, the argument proceeds as follows. The consequences of the Money Pump are unambiguously negative. It is unreasonable to have preferences which make one susceptible to such negative consequences. Having such unreasonable preferences is irrational. Therefore it is irrational to have intransitive preferences.

While the argument may be valid, it is certainly not sound. One can quite rationally be unreasonable, and quite reasonably be irrational. This variation does not provide grounds for considering intransitive preferences irrational.

³³ Davidson, D., McKinsey, J., and Suppes, P. (1955) "Outlines of a Formal Theory of Value, I" *Philosophy of Science* 22, pp. 140-160.

preferences between x, y, and z, then he would not be able to make a rational choice, defined as a choice in which the alternative chosen was not less preferred to another one also available. Therefore, his preferences would be irrational.

Putting aside any evaluation of the argument, it is clear that an agent with the relevant intransitive preferences could nonetheless make a rational choice between pairs of outcomes. Therefore the Davidson, McKinsey and Suppes justification of the transitivity condition is limited to cases in which the agent is choosing between three or more options; and this leaves out many of the cases in which the transitivity condition is generally thought to apply.

For instance, it rules out the choices in the Money Pump example itself. The Money Pump example is comprised of a series of diachronic choices over pairs of options. Further, in each choice situation, the agent is able to make a rational choice, as defined by Davidson, McKinsey and Suppes, by choosing the more preferred option. As a result, the Money Pump example does not illustrate the Davidson, McKinsey and Suppes argument for which it was offered as an example.³⁴

If the Money Pump Argument does not offer a justification of the transitivity condition based on the argument offered by Davidson, McKinsey and Suppes, it is still a poignant example. Perhaps this is why it has gained currency of its own. Indeed, there is an extended version of the Money Pump Argument which

³⁴ Interestingly, according to Davidson, McKinsey and Suppes, the inspiration for the Money Pump comes from Dr. Norman Dalkey of the Rand Corporation.

seems to reflect something of the tension inherent in having intransitive preferences.

Consider an extended version of the Money Pump Argument in which the agent allows himself to be iteratively pumped for all of his money. On its surface this seems absurd, and it may be tempting to think that the Money Pump Argument can offer a justification of the transitivity condition based on a reduction ad absurdum. However, a closer examination of a potential reductio suggests this is not the case. Consider three different ways that an agent faced with an extended Money Pump might react.

In the first scenario, the agent recognizes that he faces a Money Pump, and this affects his evaluation of the options he is presented in the following way. When presented with the opportunity to trade z_n for y_n for a little bit of money, the agent recognizes that he is not only being offered a choice between z_n and y_n , but he is also being offered a choice between being pumped or not. In other words, he is being offered a choice between z_n and avoiding a Money Pump or y_n and submitting to a Money Pump.³⁵ In the spirit of the example, it can safely be assumed that the agent would prefer to avoid the Money Pump and would thereby prefer z_n and avoiding the pump to y_n and submitting to the pump. If an agent who re-individuated the options in the face of new information in this way continued to choose in accordance with the Money Pump, then his choices would be irrational, though not because of the intransitivity of his preferences, but rather because he chose a less preferred alternative. Of course, if he made

³⁵ Note that this is quite different from Rabinowicz and McLennan's focus on so-called sophisticated choosers, who recognize that they are facing a money pump and use backwards induction to navigate the choices that are offered. This is a matter of the individuation of options.

the rational choice, the agent would not be susceptible to the extended Money Pump.

In contrast, the agent may evaluate z_n and y_n solely on their relative merits, despite the fact that he recognizes that by choosing y_n he would thereby be submitting to the Money Pump. This would clearly be absurd. But to claim that the agent's doing so is irrational is to claim that rationality requires that the agent reindividuate the options. This does not support the claim that the transitivity conditions is a requirement of rationality, but makes a further claim that rationality requires that agents individuate options in a certain way.³⁶

Finally it might be the case that the agent does not recognize that he faces a Money Pump, and as a result the individuation of options does not reflect that fact. Such an agent would plausibly continue to choose in accordance with the Money Pump. However, in this case the reductio seems to lose its force. It is not obviously absurd for the agent to choose in accordance with the Money Pump, if the fact that he is being pumped is relatively obscured. What is obvious is that such an agent would suffer negative consequences as a result of his preferences, but a Consequentialist Justification, which has already been discussed, is very different from a justification based on a reductio. Once the consequences of the agent's choices are obscured from the choice itself, it is difficult to see the basis for the reductio.

In light of this, though the Money Pump does seem to offer the basis for a reductio, the reductio seems to offer little justification for the transitivity condition.

³⁶ This suggestion, raised by Broome, has further promise and will be returned to shortly.

This leaves two issues. First, the Money Pump Argument does not seem to offer a justification of the transitivity condition. And second, there is the uncomfortable fact that there is little explanation for why the Money Pump Argument seems to be such a compelling example.

To venture a suggestion on the second point, it is possible that the Money Pump Argument is such a compelling example, because every case in which an agent suffers harm because of the intransitivity of his preferences is a broadly speaking a Money Pump type case. I.e., the Money Pump Argument is an archetypal example of all cases in which an agent suffers harm as a result of his intransitive preferences.

If the Money Pump Argument does not justify the transitivity condition, what does? For now this is a pressing open question for proponents of Preference Theories. A possible justification for the transitivity condition will be explored in more detail in the positive section of this dissertation in Chapter 4: Preferences.

Is the transitivity condition vacuous?

In two influential works³⁷, Broome raised a significant challenge for justifying a meaningful version of the transitivity condition. The object of Broome's concern was broader than the transitivity condition. His focus was on what he referred to as the Moderate Humean View (MHV), according to which no single preference can be irrational, though patterns of preferences may be. This goes

³⁷ Broome, John, "Can a Humean be moderate?", in: Frey, R G (ed), *Value, Welfare, and Morality*, Cambridge University Press : New York, 1993. And, Broome, John, *Weighing Goods*, Oxford : Cambridge, Mass. : Basil Blackwell, 1991.

beyond merely an interpretation of 'preference' which has been the concern to this point. Consider the following brief description of MHV.

Reason has a role in guiding people through life. It helps people determine and modify their preferences; the ultimate basis of which is unconstrained by rationality. While no specific preference can be irrational on its own, certain patterns of preferences can be. Some preferences can give grounds for others, such that if a person has some preferences it would be irrational for her not to have others, (by transitivity, for instance). The fundamental principle: a reason is always derived from a preference. This fundamental principle is supplemented with requirements of practical rationality. I.e., if an agent has inconsistent preferences, rationality requires her to resolve the inconsistency, but does not determine in which way this should be done. As Broome concisely expresses it, according to MHV, "You may, rationally, have any preferences, provided only that they are consistent with each other. And what consistency requires is spelled out in decision theory."³⁸

Rather than an interpretation of 'preference', MHV is perhaps better understood as a conception of practical reasoning comprised of a conception of reason, the sources of reasons, and the limits of rationality. It is perhaps one of the standard conceptions of practical reasoning used to ground the concept of preferences³⁹. While MHV will not play a significant role in the remainder of this dissertation, the features of transitivity to which Broome points, and the attending requirements for offering a meaningful condition of transitivity are not

³⁸ Broome, John, "Can a Humean be moderate?", in: Frey, R G (ed), *Value, Welfare, and Morality*, Cambridge University Press : New York, 1993. And, Broome, John, *Weighing Goods*, Oxford : Cambridge, Mass. : Basil Blackwell, 1991, pg 69.

³⁹ The relation between the concept preference and conceptions of practical reasoning will receive more attention shortly.

limited to MHV. They are features of transitivity itself and apply to conceptions of 'preference' more generally. It is these aspects of Broome's discussion that will be of interest here.

There are roughly 3 steps to Broome's argument.

- Step 1: The prospect of fine individuation poses a problem to the transitivity condition and thereby to MHV, which relies on it.
- Step 2: There are only two alternatives that can save transitivity from the problems posed by fine individuation.
- Step 3: These two alternatives are themselves inconsistent with MHV.
- Conclusion: Therefore, one can give up the transitivity condition, and thus give up MHV, or one can defend the transitivity condition by stipulating either of these two alternatives, and thereby give up MHV.

If Broome is right, the Moderate Humean faces a damning dilemma. Whichever option he chooses, he must give up MHV. Since the current discussion is interested in exploring the general feature Broome points to, and not defending MHV per se, the following discussion will be focused on the first two steps of Broome's argument.

With this general background established, it is possible to carry on with an exploration of Broome's argument.

Step 1: The prospect of fine individuation poses a problem to the transitivity condition and thereby to the MHV, which relies on it.

The problem posed by the prospect of fine individuation is most easily illustrated through the example Broome considers: Maurice, when given a choice between going to the Alps (A) and going to Rome (R), he prefers to go to Rome (R). When offered a choice between staying at home (H) and going to Rome (R), he prefers to stay at home (H). And, when offered a choice between staying home and going to the Alps (A), he prefers to going to the Alps (A).

Maurice's preferences seem to be intransitive. It appears that he prefers staying home to going to Rome, going to Rome to going to the Alps, and going to the Alps to staying home, which can be represented as follows: $H > R$, $R > A$, $A > H$. This seems to be a classic case of intransitive preferences.

However, Broome points out, that Maurice has a defense against the charge of irrationality. It may be the case that Maurice sees four alternatives, instead of three. For instance, Maurice may see staying home instead of going to the Alps as cowardly, while this does not apply to the choice between staying at home and going to Rome. The alternatives over which Maurice's preferences range can be represented as follows:

- H_r : Staying at home when going to Rome was the other alternative
- R: Going to Rome
- A: Going to the Alps
- H_a : Being cowardly by staying at home when going to the Alps was the other alternative

When the alternatives over which Maurice's preferences range are finely individuated in this way, his preferences can be represented as follows: $R > A$,

$A > H_a$, $H_r > R$. It is easy to see that represented in this way Maurice's preferences are transitive, and therefore Maurice is not guilty of an irrationality.

If it is the case that whenever an instance of apparent intransitivity is encountered, it can always be explained away by individuating the outcomes more finely, transitivity will not constrain the practical preferences an agent can have.⁴⁰ Unless, that is, there is some basis on which to limit fine individuation. Succinctly put, if finer individuation is always possible, then the transitivity condition does not constrain an agent's practical preferences and seems crucially empty.

Step 2: There are two alternatives that can save transitivity from the problems posed by fine individuation.

The transitivity condition only constrains practical preferences, has content, when it is possible to determine that some agent's preferences are intransitive, and therefore irrational. If fine individuation is always possible, then every apparently intransitive set of preferences can be understood as a transitive set of preferences over more finely individuated outcomes. To save transitivity, to restore its content, it must be possible, at times, to limit the prospect of fine individuation. To understand what this involves, return to the example at hand.

⁴⁰ This is not to say that transitivity does not have any bite whatsoever, for in the example above, transitivity would still require that Maurice have the following preference: $H_r > A$. In other words, Maurice prefers staying at home when going to Rome was the other alternative, to going to the Alps. But of course this is a choice that Maurice could never face. Preferences between such outcomes Broome calls a nonpractical preference. The current discussion will not be concerned with nonpractical preferences here.

Maurice claims to be rational by more finely individuating outcomes. Broome understands this claim to rationality to be a conjunction of two further claims. First, that H_r and H_a are different outcomes. And second, that it is rational to have a preference between them. To deny Maurice's claim to rationality, one must deny the prospect of fine individuation by denying either of these claims. The two alternatives that Broome suggests can save transitivity, the Principle of Individuation by Justifiers (PIJ) and the Rational Requirement of Indifference (RRI), do so by offering the basis for denying claims of these kinds.

PIJ is intended to offer the basis for denying claims like Maurice's first claim, that H_r and H_a are different outcomes. Succinctly put, it states that two outcomes should be individuated from one another, if and only if they differ in terms of a justifier; where justifiers can be understood to "mean a difference between two putative outcomes that makes it rational to have a preference between them."⁴¹

The basic idea is easy to grasp. Given PIJ, if H_r and H_a do not differ in terms of a justifier then Maurice's claim that H_r and H_a are different outcomes can be denied. And, if Maurice's preferences range over three outcomes (H , R , and A), rather than four (H_r , H_a , R and A) then Maurice's preferences are intransitive and he is guilty of an irrationality. That transitivity can establish that Maurice's preferences are irrational with the help of PIJ, is an example of how PIJ can help save the substantive content of the transitivity condition.

It is interesting to note that is not sufficient to establish the identity of H_r and H_a . The identity of (or indifference between) the other outcomes must also be

⁴¹ Broome, *Weighing Good*, pg 103.

established to establish the agent's preferences are intransitive. To see this, consider an example concerning sequential choices over pairs of outcomes.⁴² Tom is offered a choice between coffee (x_1) and soda (y_1), then later a choice between soda (y_2) and beer (z_2), and still later a choice between beer (z_3) and coffee (x_3). In the first situation Tom chooses coffee because he prefers coffee to soda. In the second he chooses soda because he prefers soda to beer. And in the third situation he chooses beer, because he prefers beer to coffee. It appears that Tom's preferences may be intransitive, i.e. Tom prefers coffee to soda, soda to beer and beer to coffee. What remains to establish that Tom's preferences are intransitive?

To establish that Tom's preferences are intransitive it needs to be established that the coffee that Tom chose over soda in the first choice situation, x_1 , should be treated as the same as the coffee that Tom did not choose in the third choice situation, x_3 , and so on for soda and beer.⁴³ In other words, for Tom's preferences to be intransitive, it must be the case that x_1 should be treated as the same as x_3 , that y_1 should be treated as the same as y_2 , and that z_2 should be treated as the same as z_3 .

⁴² Discussions of (in)transitive preferences are generally concerned with sequential choices between pairs of outcomes. Though Broome is not explicit about this, I presume he is concerned with sequential choices as well; not merely because it is a matter of convention, but also because Broome's arguments only apply in such situations.

To see this, consider what would be the case if we were not concerned with sequential choices between pairs. For instance, consider Tom's choice between x , y , and z , if Tom's preferences were as follows: $x > y$, $y > z$, and $z > x$. Here Tom's preferences are intransitive. Tom can choose x from the alternatives x , y and z , y from the alternatives x , y and z , or z from the alternatives x , y and z . His preferences are intransitive, and they are irrational because there is no choice that Tom can make in which he does not choose an outcome that he prefers less to another that is also available. Further, finer individuation of the outcomes will not help in this case. As such, we can safely presume that Broome's comments pertain to sequential choices between pairs of outcomes.

⁴³ The question of whether two outcomes are actually the same partially depends on the ontology of outcomes. To discuss Broome's arguments against the moderate Humean view without first settling questions regarding the nature of outcomes, a somewhat clumsily vague articulation 'treated as the same as' has been adopted in the hope that it can accommodate a variety of positions on the matter.

PIJ (and RRI) offers the basis for treating each member of these pairs in the same way, and in so doing enables one to defend a meaningful transitivity condition.⁴⁴

With this discussion of PIJ, it is possible to move on to the third step of Broome's argument.

Step 3: These two alternatives are themselves inconsistent with MHV.⁴⁵

There are two prima facie reasons to accept Broome's claim that PIJ is inconsistent with MHV. First, to the extent that PIJ itself suggests that it is irrational to have a preference between two alternatives if they do not differ in some respect, it seems to conflict with the MHV position that no specific preference is irrational.

Second, PIJ depends on justifiers, and according to Broome's definition of justifiers an agent is justified in having a preference between two outcomes 'only if the outcomes differ in some good or bad respect'.⁴⁶ This is a definition

⁴⁴ At this point it is worthwhile to comment on the fact that there has been no discussion of Broome's second alternative, RRI. There are three reasons why PIJ has been the focus to the exclusion of RRI. First, while Broome sees PIJ and RRI as being closely connected and is personally more inclined to pursue RRI, he notes that the literature generally focuses on the question of individuation, and so he follows this convention in focusing on PIJ. Here it is Broome's lead and the general convention by focusing on PIJ and the general convention which are being followed.

Second, as was discussed earlier, in order to save the transitivity requirement, either PIJ or RRI will do. Since the subsequent discussion will offer a basis for PIJ, it will be sufficient to address the challenge of fine individuation.

Third, PIJ and RRI are intimately connected. As a result, the suggested view has the resources to offer an account of PIJ or RRI. So, focusing on both seems somewhat redundant.

⁴⁵ Of course, the discussion will only focus on Broome's claim that PIJ is inconsistent with MHV.

⁴⁶ Broome, *Weighing Good*, pg 106.

of justifiers that includes substantive commitments about what preferences it is rational for an agent to have, and as such is inconsistent with MHV.

This leads to Broome's conclusion that MHV faces a damning dilemma: one can give up the transitivity condition, and thus give up MHV, or one can defend the transitivity condition by stipulating either of these two alternatives, and thereby give up MHV.

Since the current discussion is not interested in defending MHV per se, the conclusion will not be pursued further here. Nonetheless, it is worthwhile to note the force of Broome's argument.

Moreover, the challenge of fine individuation is a challenge that stems from the nature of transitivity itself. And, any successful justification of the transitivity condition should deal with the problem of fine individuation. The discussion of PIJ, and to a lesser extent RRI, suggest that Broome has presented a viable way of doing so.

Challenge Reframed

In light of this the challenge can be re-framed: Is it possible to offer a justification of the completeness and transitivity condition based on the same interpretation of 'preference' that answers the challenge posed by the prospect of fine individuation?

Stage 2: The Approach

The strategy that will be pursued to address this challenge in the remainder of this dissertation is based on three observations, each of which has already been mentioned in passing. It will be helpful to make them explicit.

1. Formal / substantive distinction: The formal treatment of choice in Preference Theories can be distinguished from the substantive theories of choice on which they rely.
2. Conception of practical reasoning, not merely interpretation: While the challenges have by and large been framed as a tension between the interpretation of 'preference' and the justification of the consistency conditions which apply to them, there is a broader context which is relevant. The justification of the consistency conditions depend in part on the relevant understanding of other concepts such as rationality, reasons, value, which comprise a conception of practical reasoning.
3. Conceptions of practical reasoning are inter-defined: Concepts such as rationality, reasons, value, beliefs and desires are often inter-defined such that taken together they comprise what will be called a conception of practical reasoning.⁴⁷ More specifically, for present purposes a conception of practical reasoning will be taken to be comprised of conceptions of the relevant concepts such as reasons, rationality, value, beliefs and desires.

Recognizing this, the current discussion will seek to enumerate conceptions of the basic concepts which make up a conception of practical reasoning that has been emerging in the moral philosophy and philosophy of action. It will then

⁴⁷ As a result, it is often un-helpful to discuss them in isolation, or without at least tacit reference to the conception of practical reasoning in which they are situated.

seek to use this conception of practical reasoning as a foundation to justify the completeness and transitivity conditions.

Formal / substantive distinction:

Preference Theories are formal in character. This is not to say that substantive commitments do not play a significant role. The discussion above should attest to that. Rather it is to say that the formal character is a central element in Preference Theories. The formal aspects of the theories can be readily identified and may be consistent with different substantive commitments.

More specifically the formal aspects of Preference Theories consist of the preference relations (defined above) and the consistency requirements or axioms that apply to it. As elements of a formal theory, they are strictly separable from substantive commitments with which they are often associated. This was intended from the outset. Debreu explains the point as follows. "Allegiance to rigor dictates the axiomatic form of the analysis where the theory, in the strict sense, is logically entirely disconnected from its interpretations...Such a dichotomy reveals all the assumptions and the logical structure of the analysis. It also makes possible immediate extensions of that analysis without modification of the theory by simply reinterpretations of concepts..."⁴⁸

This provides a great deal of flexibility to the application of the formal elements of Preference Theory, as is illustrated by Broome's use of them to examine questions in ethics.

⁴⁸ Debreu 1959 Pg x.

“I am thinking of the core theory of economics: ‘preference theory’ it is sometimes called, or ‘consumer theory’ or ‘utility theory’, or ‘expected utility theory’... Formally, the theory is nothing more than the axioms and theorems. It can be reinterpreted by substituting in place of preference any other relation that happens to satisfy the axioms. Take these two relations:

Betterness relation for a person. _ is better for the person than _.

General betterness relation. _ is better than _.

I think each of these betterness relations satisfies the axioms fairly well, though not perfectly. Each probably satisfies them better than people’s preferences do in practice. So, provided we are careful, the whole of utility theory is available to provide an analysis of the structure of betterness.”⁴⁹

The current effort is less ambitious than Broome’s. The formal treatment of choices in Preference Theories will not be directly applied to questions in ethics. But the subsequent discussion will examine the relation between these formal theories and the substantive grounds on which they depend. For present purposes, it is sufficient to note that these aspects of Preference Theories can be distinguished from the substantive commitments.

Conception of practical reasoning, not merely interpretation:

To this point the discussion has focused on the question of the interpretation of ‘preference’ and the tension with the associated consistency conditions. For example, in the succinct discussion of Revealed Preference Theory, it was noted that Samuelson interpreted ‘preference’ as choice, and Sen showed that

⁴⁹ Broome, Ethics Out of Economics, pg 8,9.

this raised issues for justifying WARP, the consistency condition on which Revealed Preference depends. And the force of Mandler's argument stemmed from the fact the interpreting 'preference' as choice offered a possible justification of the completeness condition, while the standard justification of the transitivity condition relied on interpreting 'preference' as welfare judgments. Further, this presents a challenge for Preference Theory because the two interpretations are in a sense incompatible.

There is, however, a larger context that is relevant, particularly if the completeness and/or transitivity condition are to be considered conditions of rationality. The point is straightforward. To offer a justification of conditions of rationality, a view of rationality and its connection, or possible connection, to other concepts such as preferences and choices is at least implicitly required. This became clear in the discussion of the Money Pump Argument's possible justification, or lack thereof, of the transitivity condition. It turned out that it was not only important whether or not 'preference' was interpreted as a welfare judgment, but whether the Consequentialist Justification was viable as a justification for a condition of rationality.

A similar point was brought to light in Broome's discussion of the transitivity condition. The challenge posed by the prospect of fine individuation is a result of features of the transitivity condition that any successful account of preference will likely have to meet. However, this posed a damning dilemma to MHV, because of the substantive commitments MHV has regarding the relation between preferences and rationality. The dilemma itself results from features of

MHV, not the transitivity condition. Plausibly a different substantive conception of practical reasoning could avoid this dilemma.

There are a collection of concepts related to practical reasoning, such as reasons, rationality, preferences, value, beliefs and desires. These concepts can be understood or conceived of in different ways. Taken together a given collection or set of conceptions of these concepts comprise a possible view of how agents choose, or a conception of practical reasoning.

The examples above illustrate that the justification of the consistency conditions depends on the associated conception of practical reasoning.

Conceptions of practical reasoning are inter-defined:

In a sense, the relation between the consistency conditions and associated conception of practical reasoning is not unique. The concepts that comprise a conception of practical reasoning are also inter-related.

Take for example what Broome refers to as the Preference-Satisfaction theory of good. “What is the substantive theory implicit in economics? It is not a complete theory of good, but only a theory about what is good for a person. It is specifically:

Preference-satisfaction theory of good. One thing *A* is better for a person than another thing *B* if and only if the person prefers *A* to *B*.

(Actually, this biconditional does not express the preference-satisfaction theory completely. The theory also requires the determination to go from right to left:

when a person prefers *A* to *B*, that makes it the case that *A* is better for her than *B*.)”⁵⁰

Here the good is defined in terms of an agent’s preferences. In a similar vein, Fehige and Wessels define the rational and the moral in terms of preferences. “Preferences, so the received opinion, are the alpha and omega of practical reasoning: people are *rational* if they do what they believe will best satisfy *their own* preferences (parenthetical omitted); and people are *moral* if they do what will satisfy *everybody’s* preferences (parenthetical omitted).”⁵¹

In contrast rationality may be thought to place independent limits on preferences. For example, in the earlier discussion of the challenge posed by fine individuation suggested that saving the transitivity condition may require rationality to place limits on the individual preferences agent’s can rationally have.

The possibilities are multiple⁵², and the point here is not to offer a catalogue of different approaches. Rather for present purposes it is important to note that claims about what rationality requires, for instance, or what conditions can be justified, depend in part on the relevant conception of rationality. And the relevant conception of rationality depends in part on the broader conception of practical reasoning.

⁵⁰ Broome, Ethics Out of Economics, pg 3.

⁵¹ Fehige and Wessels, Preferences, pg xxv – xxvi.

⁵² To provide a brief indication of the range of possibilities, it is possible to:

- define reasons in terms of desires as Hausman does, or desires in terms of reasons as Scanlon suggests;
- define value in terms of reason as Scanlon does or describe reasons as dependent on value, as Raz does;
- define rationality in terms of reasons, or distinguish rationality from reasons altogether.

Further, since the conception of key concepts within a conception of practical reasoning is in part derived from its context within a conception of practical reasoning, there is little purpose to argue which conception is more appropriate simpliciter – that is to say without reference to the relevant conception of practical reasoning.

To summarize, possible justifications of the completeness and transitivity conditions depend on the relevant conception of rationality. The relevant conception of rationality in turn depends on the broader conception of practical reasoning.

Strategy:

In light of these observations this dissertation will seek to offer a justification of the completeness and transitivity condition by elaborating a conception of practical reasoning that has been emerging in moral philosophy, philosophy of law and action theory. Specifically, the discussion will be comprised of three substantive chapters. Chapter 2: Reasons is based on a conception of reasons introduced by Raz in Practical Reasons and Norms⁵³. Chapter 3: Rationality will seek to elaborate on a conception of rationality that builds on this view of reasons. And Chapter 4: Preferences will offer an account of preferences situated within the context of these conceptions of reasons and rationality. From this vantage point the discussion will explore whether the suggested account can offer a justification of the completeness and transitivity conditions that addresses the challenge posed by the prospect of fine individuation.

⁵³ Raz, Practical Reasons and Norms, London : Hutchinson, 1975.

It is worthwhile to make a few comments before proceeding. First, while the following the discussion will focus on core elements of a conception of practical reasoning, it will touch on others such as value and desires, and omit still others. In this way the discussion will inevitably be incomplete. The focus of the current work is to explore a possible justification for the completeness and transitivity conditions. There is no pretense of offering an exhaustive elaboration of a conception of practical reasoning. Further, while the discussion is in some way anchored by a conception of reasons introduced to Raz, it both draws together views currently in the literature and suggests others. Effort will be made to distinguish between the two.

That said, to say the goal of the current effort is to merely offer a justification of the completeness and transitivity conditions would be misleading. There is a larger goal which the discussion seeks to serve. The fields of study mentioned above deal with choice as a substantive matter. For the moment they will be referred to as Substantive Choice Theories. Preference Theories, on the other hand, deal with choice as a primarily formal matter. While their domains of study often overlap, currently there is little connection between the two⁵⁴. This naturally raises the question of whether there is a relation between the two, and if so, what it is?

The promise of elaborating a connection between Preference Theories and Substantive Choice Theories is greater than merely answering an intriguing intellectual curiosity. As a prodigious formal theory, Preference Theories are widespread. The conclusions that are drawn from them guide individual actions

⁵⁴ Broome's work employing the formalism of Preference Theory to investigate the structure of the good may be a notable recent exception.

and national policies; and the questions to which they are applied continue to grow rapidly. However, as evidenced by the current discussion, Preference Theories at times seem to lack the resources to adequately address complex normative challenges⁵⁵. On the other hand, Substantive Choice Theories offer the basis for a nuanced treatment of the complex normative challenges which are a regular feature of agents' daily lives, but – possibly because of this – are often confined to the rarefied atmosphere of academic discourse; rarely effecting the issue which are the subject of their concern.

Understanding the relationship, if any, between Preference Theories and Substantive Choice Theories, may provide this powerful and widespread formalism with the substantive resources to more adequately address the challenges with which it is tasked, while opening new domains for substantive exploration.

It may be objected, that part of the allure of Preference Theories was precisely that they did not require such substantive commitments. To this objection it can be said, that first as was already discussed, nothing herein suggests that Preference Theories need be committed to the suggested conception of practical reasoning. They may be compatible with many others. And second, it is hoped that the previous discussion showed this aspect of Preference Theories to be more mirage than virtue.

Moreover, if the current effort can offer a justification of the completeness and transitivity conditions, it may have many other theoretical virtues to recommend

⁵⁵ It may be claimed that complex normative challenges are not the subject matter of Preference Theories. But even the current question of how to justify the completeness and transitivity conditions is a complex normative matter in its nature and in its ramifications.

it. For example, it will expand the explanatory powers of Preference Theories and show how Preference Theories fit well with other theoretical commitments.

If this is part of the motivation for undertaking this challenge, it is nonetheless acknowledged that the current effort is only a small start.

Chapter 2: Reasons

“It is generally agreed that the notion of a normative reason cannot be explained through an eliminative definition. That is, any explanation of it in which the word ‘reason’ does not occur will include another term or phrase whose meaning is close to that of ‘a reason’ so that those who puzzle over the nature of reasons will not be helped by the definition. It will raise similar puzzles in their mind. We explain the notion of a normative reason by setting out its complex inter-relations to other concepts. Not to explain, but to minimally locate what we are talking about, we can say that normative reasons, if there are such, count in favour of that for which they are reasons. They have the potential to (that is, they may) justify and require that which they favour.”⁵⁶

In seeking to enumerate a conception of practical reasoning consistent with Preference Theory this chapter will focus on the foundational concept of reasons.

Acknowledging the methodological implications of Raz’s point the discussion will eschew attempts to offer an eliminative definition and instead proceed in two stages.

The first stage will focus on the complex inter-relations between reasons and three key concepts often associated with Preference Theory, preferences, desires and value. This section will help to focus on the relevant concept of reasons and locate the discussion relative to Preference Theory. In two different ways this section is also something of a promissory note. First in that it

⁵⁶ Raz, “Reasons: Explanatory and Normative”, pg 6.

introduces aspects of the relationship between reasons and preferences, desires and values that will only be fully described in subsequent chapters. Second in that begins a broader reconsideration of commitments associated with Preference Theories that is a goal of this dissertation as a whole.

The second stage turns its attention in the opposite direction. Rather than examining the relationship between reasons and other concepts, the second stage focuses on the conception of reasons itself. This stage will seek to expand upon an analysis of reasons offered by Raz in his seminal work "Practical Reasons and Norms" by offering an approach for unpacking the concept into its constitutive elements. In so doing, this section seeks to demystify aspects of reasons talk; and provide more tools for ascertaining whether or not a fact is a reason, whether it is the same or a different reason than another fact, and what it contributes to the reasons of which it is a part.

In so far as the second stage sheds light on the structure of reasons, it will also make another contribution to the larger goal of articulating a conception of practical reasoning that can anchor Preference Theories. For Preference Theories themselves have a formal structure, and as such it will be interesting to reflect on the possible compatibility of the two, noting similarities and differences. Indeed, it will be suggested that it is fruitful to understand reasons as having a structure that is consistent with the formal structure of preferences in an interesting way.

--

At first glance the suggested view of reasons would appear to be at odds with Preference Theories. Not so much because of any obvious conflict with Preference Theories themselves, but because of conflicts with at least three views often associated with them, specifically:

- Preferences prima facie provide reasons
- Desires prima facie provide reasons
- Something is good if it gives us reasons.

Appearances, it will be argued, can be deceiving. While the suggested view is at odds with each of these claims, the cost of giving them up is not as great as one may imagine for two reasons. First, because some these claims are not as appealing as they may first appear. And second, because in each of these cases the suggested view has the resources to describe the relationship between these concepts in a way that is consistent with what one might call the broader spirit of these claims. Each of these will be discussed in turn.

Do preferences prima facie provide reasons? Discussion of this question will, by necessity, be limited at the moment as discussion of the relationship between reasons and preferences depends on much that will be addressed in the next three chapters. Even so, the intuitive appeal of the claim that preferences do prima facie provide reasons may present a powerful objection to the suggested view and it is important to acknowledge this at the outset.

The view that preferences prima facie provide reasons seems widespread, even if it is often unstated. Consider the case of Tom and Jane deciding whether to go sailing or surfing. Jane prefers sailing to surfing, and Tom is

indifferent between the two. All other things being equal, Jane's preferences seem to provide reasons for them to go sailing instead of surfing. Indeed this intuition seems profoundly scalable. It appears to apply to Jane's individual preferences, e.g., Jane's preferences seem to give her a reason to go sailing over surfing. And it also seems to apply to societies or communities as a whole. E.g., if some prefer state A over state B, and none prefer state B over state A, then there seems to be a reason for the society or community to choose state A over state B.

Contrary to this intuition, on the suggested view preferences do not *prima facie* provide reasons. Over the next two chapters it will be argued that this is a virtue and not a vice of the position. More specifically, a view of reasons and preferences will be offered in which preferences do not *prima facie* provide reasons, but that can account for the fact that in some cases Jane may be irrational for not choosing to go sailing over surfing, that Tom may have reason to go sailing because of Jane's preferences for doing so, and that the fact that some prefer A to B may be worthy of consideration when deciding between states. Further it will be suggested that these are situations of very different kinds and that some characteristics in which they differ are normatively salient. That the suggested view is able to account for these differences and arrive at conclusions that resonate with many of our intuitive judgments seems to capture the spirit of the view that preferences *prima facie* provide reasons.

Unfortunately explaining how and why the suggested view arrives at these conclusions requires significant groundwork that will begin with a discussion of the two other related but quite distinct claims rejected by the suggested view

mentioned above. For the time being, fulfilling the promissory note of explaining the relationship between reasons and preferences will have to wait.

Do desires prima facie provide reasons? The intuitive appeal of the idea that desires prima facie provide reasons may be even stronger than the view that preferences do. After all, in some cases an agent has a reason to take action because of the desires he has. For example, in some cases an agent has reason to eat vanilla ice cream because he desires vanilla ice cream. His desire to eat ice cream seems to provide him a prima facie reason to do so.

To deny that an agent's desires prima facie provide reasons seems significant. However, exploring this issue further will help put things in a different perspective⁵⁷.

First, it is important to identify the relevant concept of reasons. For example, in many cases it may be accurate to say of an agent that he ate the vanilla ice cream because he desired to do so. In cases like this, when an agent acts on a reason, the reason is operative; and can be referred to as an operative reason.

Alternatively there are reasons that an agent has to act whether or not he acts on that reasons. These reasons are considerations that count in favor of an agent acting in one way or another. These reasons indicate what the agent should or ought to do. These reasons are normative reasons.

⁵⁷ This is a well established distinction discussed among other places in Raz, Practical Reasons and Norms.

A reason may be an operative reason without being a normative reason. And a reason may be a normative reason without being an operative reason.

Normative reasons and operative reasons can be strictly separable. However, there are also times when an agent acts on his normative reason, in which case the agent's operative reason is also his normative reason. To the extent that these two different reason concepts are distinct, this chapter is concerned with normative rather than operative reasons. Cached out in these terms, the claim associated with the suggested view is that the agent's desires may give him no (normative) reason to eat the ice cream, even though it may be his (operative) reason for eating it.

There is a similar but different distinction to be drawn between normative and explanatory reasons. Explanatory reasons are answers to why questions. For example, when asked why the agent drove to the train station, it may be appropriate to answer that the agent drove to the station because he had promised to pick up a friend. In contrast, '(normative) reasons for an action are considerations which count in favour of that action'⁵⁸. Normative reasons 'guide decision(s) and action(s), and form a basis for their evaluation'⁵⁹

At first glance the fact that explanatory and normative reasons play different roles would appear to be sufficient to draw a distinction between them. But the situation is more complex. For, in that same situation, the fact that the agent promised to pick up a friend may also be his normative reason. One and the same fact, the fact that the agent promised to go to the train station, is the explanatory and normative reason.

⁵⁸ Raz, *Practical Reason and Norms*, 186. The word "normative" has been added for the sake of clarity.

⁵⁹ Raz, "Reasons: explanatory and normative", pg 1. The plural form has been added.

This may raise questions about whether there is a difference between explanatory and normative reasons. Of course, not every explanatory reason is also a normative reason. For example, when asked why the agent drove to the train station, it may also be appropriate to answer that driving is the fastest way to get to the station, or that the agent is not concerned about the environment. Both may be perfectly acceptable explanations of the agent's action. However, they do not pertain to the same normative considerations, and – absent further elaboration – are not normative reasons.

So if normative reasons are not just the same as explanatory reasons, perhaps they are a subset of explanatory reasons; specifically the explanatory reasons that explain normative considerations. In "Reasons", John Broome suggests something similar to this. More precisely Broome defines two normative senses of 'a reason'. "A perfect reason for you to Φ is defined as a fact that explains why you ought to Φ ."⁶⁰ And, "a pro tanto reason for you to Φ is a fact that plays the for- Φ role in a weighing explanation of why you ought to Φ , or in a weighing explanation of why you ought not to Φ , or in a weighing explanation of why it is not the case that you ought to Φ and not the case that you ought not to Φ ."⁶¹ For Broome, reasons, whether perfect or pro tanto reasons, explain normative facts, specifically ought facts.

Broome argues that this definition is an elaboration of the 'common place' description of normative reasons that Raz offers, in as far as he has offered a definition of what is it for a consideration to count in favor of. "To count in

⁶⁰ Broome, "Reasons", pg 6.

⁶¹ Broome, "Reasons", pg 10.

favour of Φ is to play a particular role in an explanation of why you ought to Φ .”⁶² Contrary to Raz, Broome claims that ‘counting in favor of’ cannot be the basic normative notion, “because it is complex. It incorporates the two elements of normativity and explanation. The notion of a reason has the same complexity.”⁶³ And by extension, contrary to Raz, Broome argues that reasons cannot be the basic normative notion.⁶⁴

Disagreeing with the force of Broome’s point, Raz acknowledges the explanatory and normative dimension of reasons, and elaborates on them in a way that sheds light on the complexity of the relation between them. Raz does this by distinguishing two different concepts, explanatory reasons and normative reasons. Explanatory reasons are facts which explain why something or other is the case or has happened. Explanatory reasons, as already mentioned are answers to why questions. Further as Raz says, “For every fact there may be a reason-why question, in a correct reply to which it figures.”⁶⁵ I.e., every fact may be an explanatory reason. And, to “refer to a fact as an explanatory reason is to refer, at least implicitly, to a relation it has to

⁶² Broome, “Reasons”, pg 12.

⁶³ Broome, “Reasons”, pg 12.

⁶⁴ It should be noted that Broome’s argument is more extensive than this. For example, Broome takes the concept of ought as primitive, and places an emphasis on ‘ought facts’ as a central normative concept. Further, Broome introduces normative requirements, which govern wide scope oughts, as distinct from reasons. (More reflections on normative requirements will be introduced in the next chapter.) And, Broome codifies definitions of reasons, perfect reasons and pro tanto reasons, which are quite unique. Raz, in “Reasons: explanatory and normative”, argues against each of these points. While interesting in its own, the entirety of this debate will not be reviewed in the current discussion. Instead, it will remark on the aspects of the debate that are important for identifying the concept at hand, normative reasons, and remark on some of the points of contention, such as the existence of normative requirements, when reflecting on the fruitfulness of the suggested view.

Before moving on, it may be worthwhile raising the possibility that though Broome’s argument focuses on disagreements about specific concepts within practical reasoning, taken together, it may represent an alternative conception of practical reasoning. And, in a line of argument amicable to some of his comments, the appropriate form of assessment may be the fruitfulness of the overall conception rather than debates about the specific concepts. But be that as it may, nothing in this chapter turns on whether this is the case.

⁶⁵ Raz, “Reasons: explanatory and normative”, pg 2.

something else: it is a reason why this or that happened, etc.”⁶⁶ I.e., to refer to a fact as an explanatory reason is to refer to a fact as an element within the explanation relation.

In contrast, normative reasons are facts which can justify or require, and can guide an agent's action. In Practical Reasons and Norms⁶⁷ Raz analyzes reason statements as relations between facts and persons. More specifically he represents atomic reason statements as statements that can be symbolized as follows: $R(\Phi)p,x$. Where 'p' is a fact, 'x' is a person and 'R(Φ)' is the reason to Φ operator. This seems to track common reason statements. For example as Raz writes 'p is a reason for x to Φ ' asserts that p is a reason for x to Φ . Which neatly fits the symbolism $R(\Phi)p,x$; and is true just in case p is true and it is a reason to Φ . Referring to a fact as a normative reason is to refer, at least implicitly, to a relation it has to something else. I.e., to refer to a fact as a normative reason is to refer to it as an element within the normative reason relation.

There is no difficulty or great complexity in a fact playing a role in two different relations. Reasons can be explanatory and normative. One and the same fact can be an explanatory reason and a normative reason. And there may be instances in which a fact is an explanatory reason in virtue of being a normative reason; as was the case with the agent who drove to the train station because he had promised to pick up a friend. The agent's promise to pick up his friend was both a normative and explanatory reason. And it explained the agent's

⁶⁶ Raz, "Reasons: explanatory and normative", pg 2.

⁶⁷ Raz, J (1999), Practical Reasons and Norms, Oxford : Oxford University Press.

action because it was the normative reason on which he acted.⁶⁸ Further that a fact cannot be a normative reason without possibly also being an explanatory reason does not mean that normative reasons are merely a subset of all the explanatory reasons.⁶⁹ As we have seen, normative reasons can be distinguished from explanatory reasons in part through reference to the distinct relations which obtain. To the extent that these two different reason concepts are distinct, this chapter is concerned with normative rather than explanatory reasons.⁷⁰

Having distinguished these different reason concepts, the claim at hand can be sharpened. The suggested view denies that desires *prima facie* provide normative reasons. As was seen, the suggested view has no qualms with desires as operative or explanatory reasons. However, even the more modest claim that denies that desires *prima facie* provide reasons may strike those

⁶⁸ This begins to touch on a significant topic in its own right, the relationship between reasons and explanation. Raz deals with this explicitly under the heading of the 'explanatory/normative nexus' in "Reasons: explanatory and normative", and the topic is dealt with at length by Williams, Frankfurt, Setiya, Brandom, etc.

This topic will not be addressed further here for two reasons. First, because the goal of this section is to specify the concept of reasons, normative reasons, that is of interest to the current discussion. This can be accomplished by distinguishing normative reasons from explanatory reasons without fully exploring the nature of the possible relationships between the two. Second, because a more elaborate account of the relationship between normative and explanatory reasons involves concepts such as rationality and reason that will receive further elaboration in subsequent chapters of this current work. Moving to a discussion that significantly relies on these concepts without the elaboration would be premature.

⁶⁹ For example, I am a son and a brother. To refer to me as a son is at least implicitly to refer to my relation to others, e.g., my parents. To refer to me as a brother is at least implicitly to refer to my relation to an other, e.g., my sister. In part I am a brother in virtue of being a son. And I cannot be a brother without being a son, but this is not to say that being a brother is a type of being a son, even if it is a subset of all sons who are brothers.

⁷⁰ It is important to note the scope of the current argument. This is an argument against a negative argument Broome's position could pose to the suggested view. It is not an argument against Broome's conception of ought facts as the basic normative concept and their relationship to reasons. While this is a different view than the one suggested here, it is entirely possible that it can also be consistently articulated within a broader conception of practical reasoning. And for the moment, it can be assumed that this is the case. Supporting the suggested view does not require disproving all alternatives, only answering the poignant objections they may offer. If more than one consistent conception of practical reasoning can be offered, adjudication between or amongst them may be based on such characteristics as fruitfulness.

coming from a grounding in Preference Theory as troublesome. After all, normative reasons were said to 'guide decision(s) and action(s), and form a basis for their evaluation'⁷¹, and this seems a role desires may play. Why is this not a compelling objection to the suggested view?

Reflect for a moment on what it may be for a something to be a prima facie reason. One way to cache out the thought that desires prima facie provide reasons is to understand desires as sufficient for reasons. I.e., if an agent has a desire to eat ice cream, he has a reason to eat ice cream. But this is likely too strong as it does not capture the prima facie character of the reasons desires are supposed to provide. A simple variation may be address this: desires are sufficient for reasons, if other competing or cancelling conditions do not obtain.

While some may hold the view that desires (subject to appropriate qualifications) are sufficient for reasons, this view is not as widespread as Preference Theorist may initially think. For example, Williams notably initiated one of the most prominent recent debates about reasons, the internalist/externalist debate, by arguing for an association between reasons and desires. Yet in explaining his own view, Williams denied the claim that "A has a reason to Φ iff A has some desire the satisfaction of which will be served by his Φ -ing."⁷²; and argued that this view should not be attributed to Hume either. Most proponents of the internalist position have similarly denied the claim that desires are sufficient for reasons.

⁷¹ Raz, "Reasons: explanatory and normative", pg 1. The plural form has been added.

⁷² Williams 1981, pg 101.

For Williams the connection between reasons and desires is necessary, not sufficient. And while the view that desires are a necessary condition for reasons does strain the natural understanding of the view that desires prima facie provide reasons – or more accurately speaking: desires prima facie provide reasons. However, it may be within the scope of the plausible to say that as a necessary condition, desires prima facie provide reasons in so far as if an agent has the relevant desire he has a necessary condition for there to be a reason. But the suggested view denies even this weakened and strained version of the claim that desires prima facie provide reasons.

Once again the suggested view is in good company. While some internalists may hold that having actual desires is a necessary condition for having reasons, for many the relationship is far more nuanced. For example for Williams, in order for an agent to have a reason⁷³ it must be appropriately related to an element or disposition in his subjective motivational set. If it is not immediately obvious how far removed this is from the initial starting point, consider a relevant version of desires in the context of the original example. For example, it seems far less intuitively attractive to claim that the rational possibility of being motivated to eat vanilla ice cream provides one a prima facie reason to do so.

So in summary, while it is acknowledged that the view that desire prima facie provide normative reasons may be intuitively appealing, the fact that the suggested view denies this claim should not count heavily against it since it is

⁷³ For the moment I am taking liberties with the articulation, but the point should translate. I.e., strictly speaking this is a composition of two distinct points Williams makes. First in order for an agent to have an internal reason he must have the appropriate element of disposition in S. And second, agents only have internal reasons.

not a claim that is as widely held as one may assume. Even proponents of the internalist position who argue that desires have an important relationship with reasons do not embrace this claim. And while the suggested view admittedly goes further than internalist such as Williams and denies that there is any necessary connection between reasons and desires – except in so far as this necessary connection is a stand in for the potential motivational efficacy of reasons – the fact that it denies the claim that desires prima facie provide reasons should not count significantly against its potential fruitfulness.

While this is an argument to the effect that a consequence of the suggested view, i.e., that desires do not prima facie provide normative reasons, should not weigh heavily against the suggested view, it does not explain why this is a consequence of the suggested view, or why one should accept this position. Both will be coming shortly. But first it is important to note that though the suggested view denies that desires prima facie provide reasons, it need not deny that in some instances an agent has reason to eat vanilla ice cream because of his desire to do so. Even according to the suggested view desires can be reason effecting; and, it will be suggested, this may go a long way towards accommodating the intuitive appeal to the idea that desires provide reasons. So for the moment the discussion of the relationship between reasons and desires concludes with two promissory notes: 1) why the suggested view denies that desires prima facie provide reasons and 2) an explanation of the role of the reason effecting role of desires.

Payment of these promissory notes requires discussion of the third view, i.e., that something is good if it gives us reasons. In its current articulation this claim

is perhaps the least obviously associated with Preference Theory of the three listed. That is because this articulation comes from the buck-passing debate which itself is less connected to Preference Theories. Consider instead an earlier variation of the claim in a position that has become known as FA-Analysis⁷⁴. The basic idea behind FA-Analysis is that ‘to be valuable is to be a fitting object of a pro-attitude’⁷⁵. Here ‘pro-attitude’ can be broadly understood to include desires and preferences. Understood in this way, the basic idea behind FA-Analysis is that something is valuable if it is a fitting object of an agent’s desires, preferences, or other pro-attitudes. Or, in other words, something is valuable if an agent has reason to desire or prefer it. In this form FA-Analysis, and by extension buck-passing, may seem much more closely associated with Preference Theories.

Further, according to a recent history compiled by Rabinowicz and Rønnow-Rasmussen, FA-Analysis and variations thereof, are widely accepted.

Proponents include: Brentano⁷⁶, Ewing⁷⁷, Rawls⁷⁸, McDowell⁷⁹, Chisholm⁸⁰, Wiggins⁸¹, Gibbard⁸², Anderson⁸³ and Rabinowicz⁸⁴.

⁷⁴ In recent work by Heuer and others, it has become clear that there is significant reason not to associate buck-passing too closely with FA-Analysis. There are a myriad of difficulties faced by FA-Analysis that buck-passing does not face. The contrary may also be true. At this point this should not be too great a concern. The association of the two is meant to locate a family of views associated with Preference Theories. Where relevant the challenges that each face will be addressed separately.

⁷⁵ Rabinowicz and Rønnow-Rasmussen, pg 204.

⁷⁶ Brentano, Franz (1969 [1889]), *The Origin of Our Knowledge of Right and Wrong*, London and Henley: Routledge & Kegan Paul.

⁷⁷ Ewing, A. C. (1947), *The Definition of Good*, London: MacMillan.

⁷⁸ Rawls, John (1971) *A Theory of Justice*, Cambridge, Mass.: Harvard UP.

⁷⁹ McDowell, John (1985), “Values and Secondary Qualities” In Ted Honderich, ed., *Morality and Objectivity*, London & Boston: Routledge & Kegan Paul, pp. 110-129.

⁸⁰ Chisholm, Roderick M. (1986), *Brentano and Intrinsic Value*, Cambridge: Cambridge UP.

⁸¹ Wiggins, David (1987), “A Sensible Subjectivism?”, essay V in D. Wiggins, *Needs, Values, Truth: Essays in the philosophy of value*, Oxford: Blackwell.

⁸² Gibbard, Allan (1998), “Preference and Preferability” in Christoph Fehige and Ulla Wessels (eds.), *Preferences - Perspectives in Analytical Philosophy*, Berlin: Walter de Gruyter.

⁸³ Anderson, Elisabeth (1993), *Value in Ethics and Economics*, Cambridge, Mass.: Harvard UP.

⁸⁴ Rabinowicz, W and Rønnow-Rasmussen, T (2004), “The Strike of the Demon: On Fitting Pro-Attitudes and Value” *Ethics* 114 (3)

Given the illustrious list of proponents of FA-Analysis, and its contemporary variation, i.e., buck-passing, the fact that the suggested view rejects these positions can seem to count against it; but there is reason to think this is not the case. To see this, the discussion will briefly discuss a challenge to buck-passing and then turn to arguments proponents of buck-passing offer against the alternative adopted by the suggested view.

Buck-passing only emerged in its own right when introduced by Scanlon in "What We Owe Each Other"⁸⁵. As Pekka Väyrynen⁸⁶, Mark Schroeder⁸⁷ and Roger Crisp⁸⁸ have argued, buck-passing is comprised of two different and independent claims: BPA- and BPA+; the negative argument against alternatives and the positive argument for the approach respectively.

- "(BPA-) The fact that something is good or of value is not itself a reason to respond to it favorably or to behave in certain ways with regard to it."⁸⁹
- "(BPA+) The fact that something is good or of value consists in the fact that it has some other property P which is a reason to respond to it favorably or to behave in certain ways with regard to it."⁹⁰

Since it has already been acknowledged that more than one consistent conception of practical reasoning may be offered, potential difficulties facing BPA+ will not be discussed at length. For present purposes, it can be assumed

⁸⁵ Scanlon, *What We Owe to Each Other*, Cambridge: Harvard University Press 1998.

⁸⁶ Väyrynen, P. (2006), 'Resisting the Buck-Passing Account of Value', in Russ Shafer-Landau (ed), *Oxford Studies in Metaethics*, vol 1, 295-324.

⁸⁷ Schroeder, M. (2009), 'Buck-Passers' Negative Thesis' in *Philosophical Explorations* 12(3): 341-47.

⁸⁸ Crisp, R. (2005), 'Value, Reasons, and the Structure of Justification: How to Avoid Passing the Buck' in *Analysis* 65: 80-85.

⁸⁹ Heuer, "Beyond Wrong Reasons: The Buck-Passing Account of Value" pg 1.

⁹⁰ Heuer, "Beyond Wrong Reasons: The Buck-Passing Account of Value" pg 2.

that these challenges can be overcome. Instead, one drawback will be noted because it speaks to the fruitfulness of any such account.

Taking for granted for the moment that there are versions of buck-passing that do not face the much discussed Wrong Kind of Reason challenge, there is a different challenge that may be more fundamental. In defining value as that for which there are appropriate reasons, one needs the ability to offer a non-circular description of reasons. I.e., one needs to offer a description of reasons that does not crucially rely on the role of value. This is perhaps clearest when articulated in terms of FA-Analysis. If 'to be valuable is to be a fitting object of a pro-attitude'⁹¹, what makes something a fitting object of a pro-attitude that is not itself something valuable or good? For example, if eating vanilla ice cream is valuable because it is a fitting object of a pro-attitude such as preferences or desires, what makes it a fitting object of desires or preferences? Possibly it is the fact that eating ice cream would be a refreshing desert on a hot summer day which makes it the fitting object of a desire or preference. But isn't the fact that eating the vanilla ice would be refreshing a good or valuable characteristic of eating vanilla ice cream? Of course the answer to this question depends on the relevant view of the relationship between specific evaluative qualities such as 'being refreshing' and qualities such as goodness and value. For the buck-passing account of value to work, proponent must argue that either it is not specific evaluative qualities like "being refreshing" that make eating ice cream a fitting object of preferences and desires, or they must argue that in cases like these specific evaluative properties are not good or valuable. The first disjunct seems implausible. Therefore in order to avoid circularity proponents of buck-passing or FA-Analysis are committed to a claim that

⁹¹ Rabinowicz and Rønnow-Rasmussen, pg 204.

qualities like being refreshing are not good or valuable. And indeed some seem to be committed to this claim and have used it as the basis for BPA-. As a result, the discussion of BPA- will serve to reflect on the negative argument offered by proponents of buck-passing as well as clarifying the challenge facing its proponents.

Heuer's discussion of BP- has helpfully identified three different variants that can be offered against opponents to buck-passing:

- BP1 – 'x is of value, iff it has other non-normative, natural properties that provide reasons for actions'
- BP2 – 'x is of value, iff it has either (a) other evaluative or (b) non-normative, natural properties that provide reasons for actions. There are instances of both (a) and (b), but goodness itself is not a reason.'⁹²
- BP3 – 'x is of value, iff it has other evaluative properties that provide reasons for actions, but goodness itself is not a reason.'⁹³

To begin, start with the shared thought in BP2 and BP3, namely that goodness itself does not provide reasons but other evaluative properties do. Consider the possible relationships between value or goodness and the specific evaluative properties that appear to provide reasons. Assume for the moment, as Scanlon does, that goodness is either distinct from specific evaluative properties or is a shared property of them. If the first disjunct is true, and goodness provides reasons, then it must provide a further reason than the reasons provided by specific evaluative properties, but this does not seem to be the case. To illustrate this point, Scanlon uses the example of cancer research which we

⁹² This variation has been augmented slightly.

⁹³ Heuer, pg 2.

have reason to support because it sheds light on the causes of cancer. “That the cancer research is good or valuable, Scanlon maintains, would surely not give us any *further* reason to support it in addition to the fact that it casts light on the causes of cancer.”⁹⁴ Without a role in explaining the reasons for supporting cancer research, there is reason to doubt that goodness is a distinct reason providing property.

If, on the other hand, the second disjunct is true, and goodness provides reasons, then goodness, understood here as a simple property, would appear to provide the only reason. But there are a plurality of reason providing properties that provide a plurality of reasons, so goodness cannot be a simple property shared by the specific evaluative properties that provides reasons.

If the assumed disjunction exhausts the possible relationships between goodness (or value) and specific evaluative properties, then the fact that goodness provides neither an additional reason, nor the only reason, suggests that goodness is not a reason providing property. This conclusion just is BPA-, i.e., “The fact that something is good or of value is not itself a reason to respond to it favorably or to behave in certain ways with regard to it.”⁹⁵

This is challenging, for it seems the assumptions of the argument are sound. Goodness does not provide a further reason, nor does it provide the only reason. Does this mean opponents of buck-passing must come to accept it?

Heuer offers reason to pause by questioning whether the stated disjunction exhausts the possibilities and offers an alternative understanding of the relationship between value or goodness and specific evaluative properties, the

⁹⁴ Heuer, pg 5.

⁹⁵ Heuer, “Beyond Wrong Reasons: The Buck-Passing Account of Value” pg 1.

specification relation. Where specification is understood as: "If E , F and G are variable for properties, then F specifies G iff:

- 1) necessarily all F 's are G 's, but
- 2) possibly some G 's are not F 's, and
- 3) if a G is not an F , then necessarily there is some E such that necessarily all E 's are G 's, but possibly some G 's are not E 's."⁹⁶

The suggestion is that specific evaluative properties stand in a specifying relationship to value or goodness. That is to say that specific evaluative properties specify the way in which something is of value. For example, to be delicious is a way in which food can be of value. Obviously food can be of value without being delicious, but when it is, there is some other specific property which makes it so.

Understanding this, what are the implications for Scanlon's argument? Heuer acknowledges the truth of Scanlon's assumptions: goodness does not provide a further reason nor does it provide further reasons. Yet she rejects the validity of his argument and denies the claim that therefore goodness provides no reason.

As was discussed earlier, the buck-passing account relies on a further assumption that these are the only possibilities. I.e., that if goodness provides a reason then it either

⁹⁶ Heuer, pg 20.

provides the only reason or a further reason. Since it is acknowledged that value does neither of these, and these are the only possibilities, value does not provide a reason.

The specification relation casts doubt on this assumption by showing how value can provide a reason without providing the only reason or a further reason.

Consider the following example. I have reason to read Heuer's article because doing so would be valuable. More specifically, my reason for doing so would be because it would be enlightening. And, even more specifically my reason for doing so would be because it insightfully treats an important but challenging subject with sophistication and nuance.⁹⁷

The reason that I have to read Heuer's article because doing so is valuable is not a further reason than the reason I have because doing so would be enlightening or the reason that I have because it treats an important but challenging subject with sophistication and nuance. I.e., value does not provide a further reason.

Nor is it the case that value provides the only reason. Consider an addition to the example. As in the previous version I have reason to read Heuer's article because doing so would be valuable. And I have reason to read Heuer's article because doing so would be enlightening. But in this case I have reason to read Heuer's article because it introduces important new concepts which further my understanding of reasons.

⁹⁷ This is a variation of an example Heuer offers in "Explaining Reasons" pg 22.

Treating an important but challenging subject with sophistication and nuance is quite different than introducing important new concepts. The reasons that these two facts provide are quite different. Yet each is a reason for my reading Heuer's article because doing so would be enlightening, and because doing so would be valuable. In this case, being valuable provides two different reasons. I.e., value does not provide the only reason.

It can be tempting for proponents of buck-passing to argue that these examples illustrate their point. For in these examples it is not the case that the fact that reading Heuer's article is valuable provides reasons and it is not the case that the fact that reading Heuer's article would be enlightening provide reasons. Rather it is only the most specified versions of the facts, i.e., that it treats an important and challenging subject with nuance and sophistication and that it introduces important new concepts which further my understanding, which provide reasons. But, as Heuer points out, this seems to rely on the further questionable claim that only the most specified facts provide reasons.

The claim is questionable for at least two reasons. First it seems an overly burdensome regimentation on how one thinks of reasons. For example, if the claim is accepted, then much of what are commonly considered reasons would turn out not to be. And second it is open to a continual regress. For example, the fact that Heuer's article treats an important but challenging subject with nuance and sophistication can be further specified to include characteristics such as clarity and the avoidance of pitfalls which have plagued other writers on the subject. Indeed there may be no end to such further specification as nuance and sophistication are complex evaluative properties that can be

specified in possibly infinite number of different ways. That this is the case should not lead one to conclude that I have no reason to read the Heuer's article because it treats an important but challenging subject with nuance and sophistication. So the claim that only the most specified facts provide reasons is questionable because it would introduce a costly and artificial regimentation on the understanding of reasons and possibly lead to absurd results.

Absent this claim there is no difficulty understanding how value can provide a reason without providing the only reason, nor providing a further reason. E.g., I may have many reasons to read Heuer's article because doing so would be valuable. The more specific evaluative properties specify these reasons. And at some level of specification the specific evaluative properties can distinguish one reason from the other.

By offering the specification relation, Heuer not only undermines Scanlon's negative argument, but also provides the basis for showing the internal inconsistency of BP2 and BP3. BP2 and BP3 both deny that goodness itself is a reason, yet rely on the possibility that evaluative properties provide reasons. If the specification relation holds between specific properties and goodness in general, then if evaluative properties provide reasons then goodness itself provides reasons. Since both BP2 and BP3 rely on the possibility that evaluative properties provide reasons, but deny that goodness itself is a reason, both are internally inconsistent if the specification relation holds.

Further, if the specification relation holds, then those proponents of buck-passing initially inclined to support BP2 and BP3 have reason to reject BPA- i.e., “The fact that something is good or of value is not itself a reason to respond to it favorably or to behave in certain ways with regard to it.”⁹⁸. For if specific evaluative properties provide reasons, then goodness or value provides reasons.

This leaves BP1, i.e., ‘x is of value, iff it has other non-normative, natural properties that provide reasons for actions’⁹⁹. To assess whether non-evaluative properties provide reasons, Heuer offers the Universality Requirement: “...a property P which provides a reason in a given set of circumstances C must provide the very same reason, whenever P is present in C.”¹⁰⁰

The appeal of this requirement is relatively easy to grasp. If the property in question remains the same and the circumstances remain the same, yet the reason varies, this suggests that it is not the property in question which provides the reason.

⁹⁸ Heuer, “Beyond Wrong Reasons: The Buck-Passing Account of Value” pg 1.

⁹⁹ Heuer, pg 2.

¹⁰⁰ Heuer, pg 16.

Heuer offers an inductive argument to show that non-evaluative properties do not meet the Universality Requirement.¹⁰¹ Using Heuer's example, consider whether the fact that the agent's jumping in the canal is the only way to save her life when building is on fire, provides the agent with a reason to jump in the canal. One clarificatory note. The issue at hand is whether non-evaluative properties meet the Universality Requirement; i.e., whether the same non-evaluative property provides the same reason in the same circumstances. As a result, it would muddy the waters to assume that an evaluative property was part of the circumstances. For if it is the evaluative property which provides the reason, and the circumstances (which include the evaluative property) stay the same, then the reason would also stay the same. But this would be a reflection of the role of the evaluative property in the circumstances and not of the reason providing role of the non-evaluative property. So for purposes of the example, assume the circumstances consist of non-evaluative properties such as the fact that the agent just lost her job and ended her relationship.

Consider whether the fact that jumping in the canal is the only way to save her life would provide the agent a reason in the circumstances that the agent just lost her job and ended her relationship. While it may appear that this non-evaluative fact provides the agent with a reason, leaping to this conclusion would be too quick. The fact that jumping in the canal is the only way to save her life *may* give her reason to jump, if, for instance hers was a life worth saving

¹⁰¹ "To avoid confusion, my claim is not that P cannot be a reason in C, just because it would not be a reason in other circumstances, nor that the circumstances would have to be understood as parts of the reason. The claim is that, if P is a reason to ϕ in C, it is always a reason to ϕ in C. But for any set of non-evaluative circumstances C and any non-evaluative property P, it seems to depend on further circumstances whether or not P is a reason. Thus, there is inductive evidence that no normatively significant property P in a given set of non-evaluative features of the circumstances will constitute a reason for action." – Heuer, pg 16.

now that she had been liberated from her awful job and freed from her devastating relationship. On the other hand, it *may* give her no reason at all, if for instance it was her job and relationship that gave her life meaning, and without either of them, there is little or any positive value to her life. The example may be a bit colorful, but the point should be clear enough. It may or may not be the case that the agent has reason to jump, given that she just lost her job and ended a long relationship. The non-evaluative property may or may not be a reason in one and the same circumstances. Whether or not it is a reason depends on the further evaluative property of whether or not there is positive value to her life. “The root of the problem” as Heuer says, “is that any non-evaluative consideration can be cancelled as a reason.”¹⁰² The non-evaluative fact does not meet the Universality Requirement; and therefore this non-evaluative property does not provide a reason.

Admittedly this is an inductive argument. It does not establish that non-evaluative properties do not provide reasons. But the extreme nature of the example does suggest that this is the case. If the fact that jumping in the canal is the only way to save the agent’s life does not provide her with reasons, what kind of non-evaluative facts would? So while this is not a conclusive argument against BP1, it is a compelling one.

To this point the discussion of buck-passing has covered much territory. It has provided reasons for rejecting BP1, BP2 and BP3, and shown reasons why

¹⁰² Heuer, “Explaining Reasons”, pg 16,17.

proponents of buck-passing originally inclined to the latter two should reject BPA-. However, it has not included an argument for the view that value provides reasons, but it has provided the resources to do so.

Can value or evaluative properties satisfy the Universality Requirement "...a property P which provides a reason in a given set of circumstances C must provide the very same reason, whenever P is present in C"¹⁰³? Consider Heuer's example, but this time the property P is that the agent's life is a life worth saving and the circumstances C includes non-evaluative facts such as the agent's jumping in the canal is the only way to save her life. In this case P does give the agent to jump in the canal. But, proponents of buck-passing may object, this does not parallel the exploration of whether a non-evaluative fact meets the Universality Requirement. In the previous discussion there were three sets of facts, P, C and third variable V, which alternatively took as an argument that the agent's life was worth living and that the agent's life was not worth living.

To parallel this discussion, the example can be restated as follows P: the agent's life is a life worth living, C: the building is burning and V: which can either take as arguments that jumping in the canal is the only way to save the agent's life or, for the purposes of simplicity, that jumping in the canal is the only way not to save the agent's life. In this case the agent would have a reason to jump in the canal when V took the non-evaluative fact that his was the only way

¹⁰³ Heuer, pg 16.

to save the agent's life; and the agent would not have a reason to jump in the canal when V took the non-evaluative property that jumping in the canal is the only way not to save the agent's life. The evaluative property may or may not give a reason in one and the same circumstances. Does this not show that evaluative reasons also fail to satisfy the Universality Requirement?

Obviously this question turns on what counts as the same circumstances. If the property V should be included in the circumstances, then an easy way of understanding the difference between these two cases is that the evaluative property provides different reasons in different circumstances. And in this case there is no worry about muddying the waters that recommends separating V from C. But the larger point can be made without splitting hairs about what should constitute part of the circumstances.

The fact that the agent's life is a life worth living provides the agent with reason to preserve her life. The different versions of V merely determine what actions that will involve. Importantly they do not cancel the reasons that she has. Evaluative properties are not subject to canceling conditions in the same way. As Heuer writes, "If an action is actually cruel (say) then there are no canceling conditions of there being a reason against it. If there is a canceling condition, e.g., if a cruel-looking action is in fact a necessary measure carried out in a somewhat rough way, the right conclusion to draw is not that cruelty does not

provide a reason in this case, but that the action was not cruel. If an action is cruel, its cruelty is always a reason against it.”¹⁰⁴

In other words, it is part of what it means to say that an agent’s life is a life worth living that it gives the agent reason to preserve her life. And it is part of what it means to say that an action is cruel that it gives reason not to take the action.¹⁰⁵ This suggests that, even without a detailed discussion of including V in C, at this high level of specification, evaluative properties meet the Universality Requirement. And therefore evaluative properties provide reasons. And further, if the specification relation holds, value or goodness itself provides reasons.

Returning to the initial motivation for this discussion of FA-Analysis and buck-passing it is now easy to see the basis for the suggested view’s rejection of the third claim, i.e., something is good if it gives us reasons. In contrast, on the suggested view, if something is good, it gives us reasons.

It is now also possible to make good on some of the promissory notes issued earlier. The first promissory note was to explain the relationships between reasons and preferences. At this point it is only possible to make a partial payment on this promissory note by explaining why the suggested view denies that preferences *prima facie* provide reasons (i.e., the first claim) and how

¹⁰⁴ Heuer, “Explaining Reasons”, pg 11.

¹⁰⁵ Heuer codifies this claim in the Conceptual Link: *The Conceptual Link*. The (even partial) understanding of any evaluative concept requires understanding some of the non-derivative reasons that the evaluative property that the concept refers to provides.” Heuer, “Understanding Reasons”, pg 9.

preferences may be reason effecting. It is also possible to make good on the second promissory note by explaining why the suggested view denies that desires prima facie provide reasons (i.e., the second claim) and how desires are reason effecting. Since the arguments for both are the same, they will be dealt with together.

Preference and desires are non-evaluative properties. They are mental states which may reflect an agent's evaluation of their object, but are not themselves evaluative properties. I.e., preferences and desires are not specifications of the good. As such preferences and desires do not provide reasons.

While preferences and desires do not provide reasons, they can be reason effecting. To see this, it will be illustrative to look more closely at the Universality Requirements. To restate, the Universality Requirement is: "...a property P which provides a reason in a given set of circumstances C must provide the very same reason, whenever P is present in C."¹⁰⁶

Straightforward on the surface, it can nonetheless benefit from being unpacked. To begin with, some points of clarification. First of all, according to Heuer, 'A property *provides* a reason, iff it is a reason when instantiated, e.g. the property of being cruel provides a reason iff the fact that an action is cruel is a reason (against it, in this case).'¹⁰⁷

¹⁰⁶ Heuer, pg 16.

¹⁰⁷ Heuer, pg 1.

Next, consider: 'property P ... in a given set of circumstances C.' This is perhaps most recognizable when reflecting on common reasons statements. It has long been noted that the everyday use of the term 'reason' is flexible. Indeed, statements about reasons often invoke a common body of knowledge and only point to the salient aspects of a reason. In situations of this sort one may point to some specific fact as a reason and one's interlocutors will recognize it as a reason given the circumstances. This is a familiar way to distinguish between a property and the relevant circumstances and reflects pragmatic consideration of communication situations. While the property / circumstance relationship Heuer refers to is reminiscent of this, it is also quite different. For it is not a matter of how one discusses reasons, but rather of what reasons consist. (of course there is and ought to be a relation between these, but one does not reduce to the other.)

The Universality Requirement is a regularity requirement with two conditions. The first condition calls for a property to provide a reason every time it is present in those circumstances. The second condition calls for the reason that the property provides to be the same reason every time.

As such, application of the Universality Requirement as a test of whether something provides a reason requires one be able to: 1) distinguish a property from a circumstance, 2) identify more than one property and more than one circumstance as being the same as one another, 3) identify whether something provides a reason, and 4) identify two or more reasons as being the same. Of

course given this, strictly speaking it would be circular to apply the Universality Requirement as a test of whether or not something provides a reason, because one would need to know whether it provides a reason to apply the test in the first place. But the test was evidently informative when Heuer employed it. So one question that emerges is why was it informative? What is the significance of the Universality Requirement?

While Heuer uses the Universality Requirement to argue that non-evaluative properties do not provide reasons, an argument to that effect could have relied on the definition of 'provides a reason' itself. According to the definition discussed earlier, being a reason when instantiated is a necessary and sufficient condition for a property to provide a reason. If the property can be instantiated without being a reason, then it does not provide a reason. Since the property of being the only way to save Mary's life could be instantiated without being a reason for Mary to jump in the canal, as would be the case if Mary's life was without value, being the only way to save Mary's life does not provide a reason.

Rather the significance of the Universality Requirement for present purposes is the extent to which it reflects the role of non-evaluative properties in determining the reasons that evaluative properties provide. To see this, it will be helpful to take a step back and begin by discussing the fact that evaluative properties provide reasons.

The fact that Mary's life is a life worth living is a reason for Mary to live her life. It is part of the nature of being valuable that there are reasons to realize that value. Value, as Raz argues, provides a general reason to realize or preserve that value which applies generally.

As being a life worth living stands in a specification relation to being valuable, so too does the reason to jump in the canal stand in something like a specification relation to the reason to realize or preserve the value of that life. It is likely uncontroversial to say that the facts, or non-evaluative properties that pertain determine the reason which the value of her life provides. In other words, it is the specific circumstances that Mary is in that make it the case that the reason she has is a reason to jump into the canal, because so doing will realize or preserve the value of her worthwhile life. The circumstances Mary is in determine how and what it would be to realize or preserve the value of her life, and in so doing determine of what Mary's reason consists.

This discussion makes it relatively straightforward to see the basis for the two regularity conditions of the Universality Requirement. First, evaluative properties or facts provide a reason in each circumstance, because so doing is part of the nature of value. Second, the evaluative property provides the same reason every time it pertains in the same circumstances because the same circumstances determine that taking the same action will realize or preserve the value. If it does not, then the circumstances are not the same.

This also sheds light on a reason effecting role of desires and preferences.

Consider a variation of the desire example discussed earlier. Imagine the agent desires vanilla ice cream, and having vanilla ice cream would provide him with a sense of satisfaction. However, the agent does not have a taste for chocolate ice cream, does not desire to eat chocolate ice cream and would not be satisfied if he did. It seems the agent has a reason to eat vanilla ice cream, but not chocolate, precisely because of his desire to do so.

This can seem paradoxical. If desires do not provide reasons, how is it that the agent has a reason to eat vanilla ice cream because of his desire to do so?

The explanation is straightforward. It is not the agent's desire to eat ice that provides his reason for doing so but the satisfaction he will have. The agent's desire to eat vanilla ice cream effects the reason that he has. In this case the agent's desire is directly analogous to the fact in that jumping into the canal was the only way to save the agent's life. In that case too it was the fact that the agent's life was a worth living that provided the reason. The non-evaluative property determined what action was called for by the reason.

This example naturally prompts the question: does the satisfaction of every desire provide a reason? If so, it would seem too much was made of denying that desires provide reasons when the second claim could have been re-articulated in terms of desire satisfaction. But the satisfaction of every desire does not lead to satisfaction. Or put more broadly, it is not the case that the satisfaction of every desire is valuable. The satisfaction of some desires, for example desires for an addictive drug, are not valuable. Some desires can never be satisfied. And the value of some desires is in the having of them, not

in their satisfaction; as in 'every man's reach should exceed his grasp'. So while in some cases an agent's desires determine what actions his reasons recommend, they do not provide reasons.

While denying the claim that desires *prima facie* provide reasons, the fact that the suggested view can account for this common sense example of desires affecting the reasons an agent has, goes a long way towards accommodating the intuitive appeal of the idea that desires provide reasons.

This also provides the basis for answering a natural question about the current discussion; and it is worth taking some time to address the point explicitly. As was already noted, Preference Theories themselves are not committed to any specific conception of reasons. That said, it is not uncommon to associate Preference Theories with internalist conceptions of reasons such as the view of reasons suggested by Williams. This dissertation takes a different tack, and it is natural to question the motivation for doing so. There are essentially three reasons for this.

First, as discussed in the Chapter 1: Introduction, this dissertation seeks to offer a positive argument that illustrates the potential of situating Preference Theories within a richer conception of practical reasoning. It does not argue that this is the only conception of practical reasoning against which Preference Theories could be situated. On the contrary, it was explicitly acknowledged that other conceptions of practical reasoning may also be consistent with Preference Theories. So no claim is being offered here that Preference Theories cannot be situated against a background conception of practical reasoning that relies on

an internalist conception of reasons, but only that this is not the focus of the current discussion.

Second, as the proceeding discussion shows, the internalist conception of reasons is not as congenial to Preference Theories as is often thought. To make the point more explicit, it can be helpful to briefly walk through the reason for this in detail.

It is common in decision theory to speak of the agent having a reason to choose an alternative that will lead to the most desirable outcome with the highest likelihood. Implicit in discussions of this sort is the view that the agent has a reason to choose the more desirable outcome. Or, to sharpen the point, the agent has reason because he desires the outcome. The agent's desire for the outcome provides the agent with a reason to choose the alternative that will lead to that outcome. In discussions of this sort, there is a strong connection between the agent's desires and the reasons that he has. And the nature of that relationship is that the agent's desires provide the reasons. The agent's desires are sufficient condition to provide reasons.¹⁰⁸

Reasons internalism is also committed to the view that there is a strong relationship between desires and reasons. As such, there is often a presumption that reasons internalism is congenial to decision theory.

Unfortunately, the nature of the relationship between desires and reasons that

¹⁰⁸ As has been discussed at length earlier, it is important to note that this view of the relationship between desires and reasons is not a necessary aspect of decision theory. It seems to have emerged by historical tradition. Many important figures in decision theory, e.g., Davidson, readily rely on a view of folk psychology in which desires play a central role, and are intimately connected to the explanation of operative reasons. Unfortunately, too little notice has been made of the difference between operative and normative reasons, and seemingly as a result the limitations of folk psychology to contribute to normative reasons is sometimes not appreciated.

reasons internalism is committed to is quite different than the relationship presumed in discussions of decision theory. Whereas the latter often presumes desires are sufficient (or would be jointly sufficient with the appropriate belief) to provide reasons, the former often denies this claim and instead is committed to the view that desires are a necessary condition for reasons.

The reason most reasons internalists deny the view that desire are sufficient to provide reasons is that this view leads to a pernicious conclusion. Specifically, an agent could have a reason to do bad simply because he has a desire to do so. Avoiding the pernicious consequence, most reasons internalists do not claim that desires are sufficient to provide reasons.

While it is common to think of decision theory as having a commitment to the relationship between desires and reasons, the relationship is quite different than the one to which reasons internalists are committed. As a result, reasons internalism is far less congenial to decision theory than is often thought.

And, third, variations of reasons internalism that significantly differ from the suggested view have a different unattractive consequence – what can be called a perverse consequence. This point requires a little additional background on the internalist views.

In its most basic form, reasons internalism amounts to a commitment to the claim that for all reasons there is a necessary relationship between the reason and elements of the agent's subjective motivational set. One can map a range of different internalist views by varying the nature of the relationship and the

understanding of the relevant aspects of the agent's subjective motivational set. Even a brief reflection on the different possibilities demonstrates the broad range of possible internalist positions.

There are a number of different ways to categorize internalist positions. For present purposes, it is sufficient to distinguish between meaningful and motivational internalism. To understand the distinction at hand, consider the universe of possible reasons that may pertain to a given agent. According to some internalist positions, a reason which would otherwise pertain to a given agent, will not, because of facts about the agent's subjective motivational set, or because of the nature of the relation between the agent's subjective motivational set and the reason. Consider a simple example of such an internalist position according to which in order for an agent to have a reason, it is necessary that the agent have an actual desire that would be furthered by acting in accordance with the reason. If the agent did not have an actual desire that would be furthered by acting in accordance with the reason, then the reason would not pertain to him. For example, if the agent did not have an actual desire that would be served by refraining from murdering the helpless victim, then the agent would have no reason to refrain from murdering the innocent victim.

For present purposes, internalist views like this one, which hold that the contents of the agent's motivational set meaningfully constrain the reasons that pertain to the agent, can be considered meaningful internalist views.

In contrast, other internalist views do not hold that the contents of the agent's motivational set meaningfully constrain the reasons that pertain to the agent. For example, consider a possible internalist position which holds that in order for an agent to have a reason, it is necessary that were the agent fully informed, perfectly rational and reasonable that he could form a motivation that would be furthered by acting in accordance with the reason. Presumably, such a fully informed, perfectly rational and reasonable agent could form a motivation that would be furthered by acting in accordance with any reason he may have. For example, since such an agent could form a desire that would be furthered by refraining from murdering the helpless victim, he would have reason to refrain from murdering the innocent victim.

On internalist views of this kind, the contents of the agent's subjective motivational set do not meaningfully constrain the reasons that pertain to the agent. Rather, the connection to possible motivational sets stands in for the potential motivational efficaciousness of reasons that pertain to an agent. For this reason, in the current discussion, views of these kinds will be referred to as motivational internalism.

As the discussion above illustrates, meaningful internalism faces a challenge that motivational internalism does not. Meaningful internalist views yield a perverse consequence in which agents that are sufficiently bad, for example, may have no reason to refrain from bad actions as a direct consequence of how bad they are.

While this perverse consequence is not in itself an argument against meaningful internalism it is an unattractive characteristic of views of these kinds and offers support to the decision to focus on other accounts of reasons.

Further, to the extent that motivational internalism merely stands in for the potential motivational efficaciousness of reasons, even views that are often considered externalist, like the conception of reasons focused on in this chapter, can be considered motivational internalists.

That said, there are three things to be said about the fact that this chapter does not focus on a meaningful internalist account of reasons. First, it is acknowledged that this chapter could have focused on an internalist account of reasons; no argument was offered to suggest that this was not a possibility. Second, despite the common association between internalist view of reasons and preferences theories, internalist views of reasons are not nearly as congenial to preference theories as is often assumed. And third, meaningful accounts of reasons internalism yield a perverse consequence that recommends exploring other accounts. And the view under consideration is broadly consistent with motivational accounts of reasons internalism.

If this helps clarify the nature of the relationship between desires and reasons on the suggested view, it is also worthwhile pointing out that on the suggested view preferences can be reasons in similar ways. Consider a variation of the earlier example in which Jane prefers sailing to surfing, and Tom is indifferent between the two. All other things being equal, Jane's preferences seem to provide reasons for them to go sailing instead of surfing. When deciding

between surfing and sailing, they have always chosen to go sailing because Tom was indifferent between the two so, he thought, they might as well go sailing. In this instance, they may have reason to go surfing precisely because it is Jane's preference. To establish an equal healthy partnership, it is important for Jane's preferences to sometimes carry the day.

Here, as in the previous example, Tom and Jane have a reason to go surfing because of Jane's preference for surfing, though her preference does not provide the reason. It is the value of having a healthy partnership that provides the reason, and it is the non-evaluative fact that Jane prefers surfing that determines what action that reason recommends.

While denying the claim that preferences *prima facie* provide reasons, the fact that the suggested view can account for the fact that Tom has a reason to go surfing because Jane prefers it to sailing resonates with intuitive judgments on the matter.

This first stage of this chapter sought to focus on the relevant concept of reasons and locate the discussion relative to Preference theory by focusing on the complex inter-relations between reasons and three concepts often associated with Preference Theory.

It acknowledged what may appear to be a challenge for the goal of offering a conception of practical reasoning based on the suggested view that can be

consistent with Preference Theory, i.e., the fact that the suggested view denies the following three claims that are often associated with Preference Theories.

1. Preferences prima facie provide reasons
2. Desires prima facie provide reasons
3. Something is of value if it gives us reasons

For each claim, it was argued that the challenge may seem larger than it is, for even if it denied the claims in question, it could accommodate the intuitive appeal that recommends each of them. For example the suggested view has the resources to explain an agent's preferences can be a reason, how desires can effect the reasons an agent has, and the intimate relationship between reasons and values.

In the course of doing so, the first stage also focused attention on the relevant concept of reasons, distinguishing normative reasons from the closely related concepts of operative and explanatory reasons.

This discussion also afforded the opportunity to address several negative arguments the suggested view will face. For example, the claim that reasons cannot be the basic normative notion because it incorporates the normative and the explanatory was responded to by suggesting that the normative and explanatory reason relations are distinct. And while one fact can serve as a relata in both a normative and explanatory reason relation, this need not create

any difficulties in understanding normative reasons as the basic normative notion.

This stage also discussed the challenge posed to the suggested view by FA-Analysis and buck-passing, namely the claim that value or goodness does not provide reasons. Heuer's articulation of the specification relation was used to argue for the view that if evaluative properties provide reasons, goodness or value itself provides reasons. And employing Heuer's Universality Requirement it was argued that both evaluative properties and goodness itself or value provide reasons, while non-evaluative properties do not.

Since preferences and desires are mental states and not evaluative facts, this latter claim provided the basis to explain the suggested view's denial of the claims associated with Preference Theory, that desires and preferences provide reasons. However, an analogy with the specification relation was used to suggest that while preferences and desires do not provide reasons they can effect the reasons an agent has. This, in turn, explained how the suggested view could accommodate the intuitions that recommend each of the three Preference Theory associated claims it denied without accepting the claims themselves.

While the first stage helped identify the relevant concept of reasons and explored its relationship to other central concepts in practical reasoning, it did

not explore the conception of reasons itself. That is the subject matter that will be taken up in the second stage of this chapter.

The world is normatively complex. Situated within relationships and societal constructs, living amongst institutions, traditions, cultural norms and practices, individuals face an array of choices that will realize values of different kinds in different ways. There is little wonder that reasons, as the basic normative concept, are remarkably sophisticated. Reasons compete and combine, they can reinforce one another and conflict, and they can even cancel one another. Unfortunately, in exploring this variety of complexity, discussions of reasons can be somewhat opaque, impeding understanding and progress in applying this powerful conceptual apparatus to important new domains. This, in part, is a challenge taken up by the current work in seeking to explore whether Preference Theories can be situated within a conception of practical reasoning which embraces reasons as the basic normative concept.

As a result, it behooves the current effort to try to shed light on aspects of reasons that have at times been elusive. If this first stage helped to establish that evaluative properties provide reasons, it did not say much about the reasons that they provide. For example, sometimes reasons statements include reference to evaluative properties, but often they do not. Many reason statement simply point to non-evaluative properties. E.g., the fact that James's train is arriving is a reason for John to go to the station. What role do non-

evaluative facts play in reasons? When is a fact a reason? When do two facts refer to the same reason, and when do they refer to two separate reasons?

The current stage will seek to provide a principled basis to answer these and other questions by focusing on the concept of reasons itself. In particular this stage will focus on the concept of a complete reason. Reason statements often include reference to individual facts, e.g., the fact that it is noon is a reason for John to go to the station. But these facts in isolation are not reasons. Their status as reasons depends on their relation to other facts. A complete reason would be a relation that takes as arguments each of the facts that are required for the facts to be reasons. The complete reason relation would explain why the facts are reasons. As such this stage will focus on offering an analysis of complete reasons that is consistent with the discussion to this point, but goes beyond it to incorporate non-evaluative elements. In the course of doing so, it will wrestle with the dual complexities of reason statement and reasons themselves.

To set expectations, it is important to note that this discussion will not pursue an exploration of the nature of reasons, nor make much progress in describing how reasons interact. Both are worthwhile subjects. The former is already the subject of considerable attention in the literature, and the latter is no doubt worthy of similar. However, to make progress towards the current aim the focus of this discussion will be narrowly prescribed.

In Practical Reasons and Norms¹⁰⁹ Raz analyzes reason statements as relations between facts and persons. More specifically he represents complete reason statements as statements that can be symbolized as follows: $R(\Phi)p,x$. Where 'p' is a fact, 'x' is a person and 'R(Φ)' is the reason to Φ operator. This seems to track common reason statements. For example as Raz says 'p is a reason for x to Φ ' asserts that p is a reason for x to Φ . Which neatly fits the symbolism $R(\Phi)p,x$; and is true just in case p is true and it is a reason for x to Φ . While this analysis brings to the surface the fact that reason statements like the above depend on p, they obscure what is required for p to be a reason to Φ , which seems a pertinent desiderata for an analysis of reason statements.

What became clear in the earlier discussion was that a number of different facts may play a role in a given reason. For example, while evaluative properties, p, provide reasons, it is the non-evaluative properties incorporated in the circumstances, c, which determine reasons.¹¹⁰ This suggests a variation of Raz's formalism: $R'(\Phi)p,c,x$. In other words, an evaluative fact p, provides a reason for x to Φ in c.

Unfortunately R' still obscures what it is about the relations between Φ , p, c and x that make it the case that the reason relation obtains. For example R' would take the following arguments: Φ : jump in the canal, p: the agent's life is a life worth living, c: jumping in the canal is the only chance to save the agent's life, and x: the agent. For the fact that the agent's life is a life worth living and

¹⁰⁹ Raz, J (1999), *Practical Reasons and Norms*, Oxford : Oxford University Press.

¹¹⁰ Here properties are taken to be facts.

jumping in the canal is the only chance to save the agent's life is a reason for the agent to jump into the canal. Though R' would not take the following arguments: Φ : jump in the canal, p : the agent's life is a life worth living, c : jumping in the canal is the only chance to not save the agent's life, and x : the agent. For the fact that the agent's life is a life worth living and jumping in the canal is the only chance to not save the agent's life is not a reason for the agent to jump in the river. On the contrary it is a reason for the agent not to jump in the river because doing so would destroy something worth preserving.

It would be useful if the analysis of the reasons relation made explicit the required relationship between its arguments. Specifically, the aspect that appears to be missing is the fact that the agent's taking the action will somehow realize or preserve the value of the evaluative property. This may be articulated as : The reason relation obtains if an agent's, x 's, taking an action, a , brings about an outcome or increases the likelihood of, o , which realizes or preserves some value, v . Using another variation of Raz's formalism, one can represent this as $R''(a),o,v,x$; or perhaps as $R''(a),o_v,x$. Returning to the same example R'' would take the following arguments: a : jump in the canal, o_v : the preservation of x 's life which is a life worth living, x : the agent. For the fact that x 's jumping into the canal would preserve x 's life which is a life worth living is a reason for x to jump into the canal.

As stated, the R'' relation imposes the following conditions on its arguments.

The reason is a reason for the agent, x , to take the action, a . The action must

bring about, or at least make more likely, the outcome, o. The valuable property, v, is realized in the outcome, o.¹¹¹

While there are doubtless other ways to formulate R", this formulation, together with its conditions, wears the logic of the reason relation on its sleeve and makes explicit what is required to be argument of the reason relation. Thus meeting the desiderata of a reason relation mentioned earlier.

A tension may naturally arise with the earlier description of reasons in which it was discussed that evaluative properties, p, provide reasons and circumstances, c, determine the reasons that they are. How does R" relate to this description?

For one thing, the R" and the previous description of reasons differ in tense. Whereas the p and c were discussed in the present tense, the outcome and value that R" admits are future tensed, broadly speaking. I.e., the valuable outcome that the action will bring about.

The earlier description bears revisiting. Recall the first interpretation of the example. P: the agent's life is a life worth saving. On further reflection, it is perhaps misleading to describe P as present tensed. Whether or not the agent's life is a life worth saving must depend in part on the character and quality of the agent's life after it is saved. For example, if the agent's life after being saved would not be worthwhile, it is difficult to see why the fact that her

¹¹¹ NB: Though the outcome in the example preserves the valuable life of the agent, it is not a condition of the reason relation that the valuable outcome must pertain to the agent undertaking the action. Plausibly an agent may have reason to take an action that preserves or realize value for another.

life is now worthwhile is a reason to save it. It may, on the other hand, be a reason to take other actions that would prevent the agent's life from being a life in need of saving. But this too would depend on the action's effect on value. Taking another action that prevents the agent's life from being a life in need of saving could be a reason precisely because it would preserve, in the future, the worthwhile character of the agent's life. In contrast, even if the agent's life is not now a worthwhile life, the fact that it would be worthwhile after having been saved is a reason to save the agent's life. The truth of P relies on what will happen in the future.¹¹²

At the heart of this is the fact that it is the action's effect on value that is relevant to whether or not there is a reason to take the action. And in the normal course of things, excluding cases at the limit where the temporal location of the action and the outcome become indistinguishable, this effect can be aptly described in the future tense.

In this example the temporal signatures in C are more readily transparent.

Recall that C includes non-evaluative facts such as that the agent's jumping in the canal is the only way to save her life. Here the fact itself describes events which are temporally elongated. It includes a description of the action, jumping in the canal, and the outcome, saving of the agent's life, which is the causal consequence of the action.

¹¹² Indeed this future orientation is reflected in Raz's articulation of the most general reason that value provides. Recall that Raz suggests value provides a general reason to realize or preserve that value which applies generally. The reasons to realize and preserve value are both future oriented.

Despite appearances, it seems, R" and the previous description of reasons incorporate the same elements. The apparent differences may only be a result of their different purpose. Whereas the previous descriptions of reasons sought to clarify which element provides the reasons, R" seeks to elaborate upon the relationship between the elements of a reason.

It is worth noting that R" represents reasons as broadly speaking consequentialist. That is to say that it is the value of the outcome or consequence that provides the reason. There are at least three difficulties with this. First, there are actions that are valuable in themselves. Second, in some cases the character of the action is relevant to the reason. Third, it is at least possible, and highly likely, that deontological considerations also provide reasons. Each of these would seem to be at tension with a consequentialist formalism for reasons. Each will be looked at in turn.

There are some actions which are valuable in themselves. For example, standing up for the welfare of a vulnerable group, may be valuable in itself even if doing so creates detrimental effects for oneself or others. Admittedly the detrimental effects to oneself or others may be reason not to stand up for the vulnerable group, but this is a further reason and does not cancel the reason that there is for doing so. Examples like this would seem to be a poor fit for R". Rather it is tempting to offer a variation of the formalism to acknowledge the fact that the value in these cases accrues to the action. R"', R"' (a_v), x, is a candidate for doing so.

While this is certainly a possibility, it will not be the alternative pursued here. Instead it will be suggested that R" be applied broadly enough to accommodate case of this kind. This can simply be accomplished by permitting the outcome, o, to take a description of the action, a, as an argument. Note, this is not quite the same as saying that a and o can take the same argument. For o includes an evaluative element, whereas a does not. This is not as much of a stretch as it may first appear. Consider the example in question where R" would take the following arguments: a: stand up for the welfare of a vulnerable population, o_v: stand up for the welfare of a vulnerable group which would be valuable in its own right, x: the agent. Using the plain language description of R" offered earlier this can be articulated as: the agent has a reason to stand up for the welfare of a vulnerable group if doing so would bring about or increase the likelihood of the agent's standing up for the valuable group which would be valuable in its own right. Or, less redundantly, the agent has a reason to stand up for the welfare of the valuable group if doing so would be valuable in its own right. In this way, by permitting the outcome, o, to take as an argument a description of the action which incorporates the evaluative element, R" can accommodate cases in which an action is valuable in and of itself.

This is somewhat different from cases in which the character of the action itself is relevant to the reason. For example, consider the case of Lance Armstrong. Plausibly, as a cyclist Lance Armstrong had reason to win the Tour de France, which is regarded as the most difficult cycling race in the world and the winning of which would indicate he had become the best in the world at his chosen profession. For arguably one has reason to become the best in the world at one's chosen profession, particularly if it is a competitive sport like cycling. Yet

the way in which Lance Armstrong wins the Tour de France, the character of the action, is relevant to the reason he had. For Lance Armstrong did not have a reason to win the Tour de France by taking steroids and cheating; he had reason to win the Tour de France through honest effort. How is this to be understood in terms of R"?

Rendered in the formalism, R" would take the following arguments: a: win the Tour de France by cheating, o_v : indicate Lance Armstrong had become the best in the world at his chosen profession which would be valuable, x: Lance Armstrong. Using the plain language description of R" this would yield: Lance Armstrong has a reason to win the Tour de France by cheating if doing so would bring about or increase the likelihood of indicating that Lance Armstrong had become the best in the world at his chosen profession, which would be valuable. But it is precisely because winning the Tour de France by cheating would not indicate that Lance Armstrong had become the best in the world at his chosen profession that this is not a reason for Lance Armstrong to win the Tour de France by cheating. Whereas, he did have reason to win the Tour de France through honest effort. In cases of this kind, the character of the action is relevant to the reason in that it effects whether the action can bring about the outcome or make it more likely that it comes about.

The issue with deontological consideration is quite different. There likely are deontological considerations which provide reasons. The prohibition on murder is an obvious candidate. And in any event, the formalism should not pre-judge the matter. It should be acknowledged that the current analysis of reasons pertains to value based reasons. If deontological considerations provide

reasons, the suggested view should be augmented to incorporate them.

However, this is a large topic in its own right and will not be addressed here.

While a limitation of the current account, it is not an overwhelming drawback as many of the reasons agents have are value based.

This limitation notwithstanding, for the domain to which it pertains, the current analysis and the associated conditions provide an account of complete value based reasons that wears its logic on its sleeve. E.g.,

- Formal analysis: $R''(a), o_v, x$, where R'' is the reasons relation, a is an action, o_v a valuable outcome, and x is an agent.
- Conditions:
 - The reason is a reason for the agent, x , to take the action, a .
 - The action must bring about, or at least increase the likelihood of the outcome, o .
 - The valuable property, v , is realized in the outcome, o .
- Plain language description: The reason relation obtains if the agent's, x 's, taking an action, a , brings about or increases the likelihood of a valuable outcome, o_v .

While it should be clear from methodological statements to this effect throughout this chapter, no attempt has been made to argue that this is *the* way of describing complete reasons, the elements that comprise them, and the relation that holds amongst the elements. Rather it is being suggested that this is a fruitful way for doing so. And this approach meets at least two desiderata for such an effort. It makes clear the conditions for a fact to qualify as an

element of a complete reason. It lays out the relation between the elements. And it brings to the fore the force of the reason. It is a further question whether the analysis is useful in understanding reasons statements. And that is the subject which will next be considered.

To begin with, this treatment of complete reasons can provide a systematic basis for dealing with some preliminary questions about reasons that stem from the pragmatic way that they are discussed. As was already noted, in making statements about reasons, it is common practice to mention some pertinent facts, but not elaborate the complete reason itself. While usually sufficient for the purposes at hand, this can and does lead to questions of whether two statements about facts refer to the same or different reasons. A brief discussion of an example Raz introduces can be instructive: “When asked why he goes to the station, John may say that (a) James will be arriving there, or that (b) James will be pleased to be met at the station, or that (c) he would like to please James... and that then John says that (d) he has promised James that he will meet him at the station, that (e) one ought to keep one’s promises, and that (f) one ought to please one’s friends.”¹¹³

By way of explanation, Raz says, “(a) to (c) state parts of a reason which John has for going to the station. (d) states not a further part of the same reason but a part of a second reason for the same action. (a) may well be a statement of part of the second reason as well, but not (b) or (c). (e) and (f) are quite

¹¹³ Raz, *Practical Reasons and Norms* pg 22,23

different again. They are not parts of the reasons which John has for going to the station. They explain in two different ways why John has or regards himself as having two reasons for going to the station.”¹¹⁴

In many ways, Raz’s comments are straightforward and reflect the discussion to date. Nonetheless, it will be helpful to elaborate briefly. For example, as Raz says (a) through (c) may be one reason for John to go to the station. But it is a very different reason than the reason John has to in virtue of making James happy. Relying on the recent discussion of complete reasons can make this clearer. The first reading can be rearticulated as: John has a reason to go to the station because by going to the station he will make James happy and satisfy John’s desires to make James happy.¹¹⁵ The valuable outcome that provides this reason is the value that comes from satisfying the desire to make James happy. Whereas the second reading can be rearticulated as: John has a reason to go to the station because by going to the station he will make James happy. The valuable outcome that provides this reason is James’ being happy. In the first reason, James’ happiness is instrumental to satisfying John’s desire, whereas in the second reason it is the end which provides the reason.

Pressing on, what then of the question raised earlier, whether statements referring to two facts refer to the same or different reasons. Recall that earlier it was noted that two different fact statements may refer to one and the same

¹¹⁴ Raz, pg 23.

¹¹⁵ Recall the earlier discussion about the difference between satisfying a desire and the value which can result from satisfying a desire. Strictly speaking it is the latter which is relevant here. But, for present purposes Raz’s articulation will be followed to parallel the example.

reason and that a single fact statement can be used to refer to more than one reason. To elaborate on how this was possible it was shown that a single fact may stand in two different complete reason relations and that two different facts can stand in one and the same complete reason relation. This may suggest that the complete reason relation can serve as the basis for identifying and distinguishing reasons. Perhaps through what may be called the Complete Reason Identity Relation (CRIR): the complete reason $R''(a_1, o_{v1}, x_1)$ is identical with the complete reason $R''(a_2, o_{v2}, x_2)$ if and only if $a_1 \equiv a_2$, $o_{v1} \equiv o_{v2}$, and $x_1 \equiv x_2$. However, because the CRIR depends on the specification of the elements of complete reasons and the identity relations that hold amongst them, resting with CRIR without further elaboration would be too quick.

Consider the following example. Mary has reason to go to the symphony if Mary's attending the symphony would bring about or increase the likelihood of Mary's engaging with the artistry of the music. As a result of the specification relation discussed earlier, one can regiment this reason in the following way: Mary has reason to go to the symphony, if attending the symphony would be valuable. This reason statement includes references to each of the elements of the complete reason, i.e., the agent (x_1): Mary, the action (a_1): going to the symphony, and a valuable outcome (o_{v1}): attending the symphony would be valuable. However, it is altogether possible that Mary may also have another reason for going to the symphony, for example, to appropriately engage with the artistry of the conductor which would also be valuable. One can regiment this reason statement in the following way: Mary has reason to go to the symphony if Mary's attending the symphony would bring about or increase the

likelihood of Mary's engaging with the artistry of the conductor. Once again, because of the specification relation, one can regiment this reason statement in the following way: Mary has reason to go to the symphony, if attending the symphony would be valuable. This reason statement includes references to each of the elements of the complete reason, i.e., the agent (x_2): Mary, the action (a_2): going to the symphony, and a valuable outcome (o_{v2}): attending the symphony would be valuable. It is the case that: $a_1 \equiv a_2$, $o_{v1} \equiv o_{v2}$, and $x_1 \equiv x_2$. Yet the reason statements refer to two different reasons. Put more succinctly, even complete reason relation can be ambiguous if not articulated at the appropriate level of specification. When this is the case, identity amongst the elements of complete reasons relation is not sufficient to establish that two reason statements refer to the same reason. This will be referred to as the false positive challenge, because CRIR yields false positive result that the complete reason $R''(a_1), o_{v1}, x_1$ is identical with the *complete reason* $R''(a_2), o_{v2}, x_2$.

Consider another variation of an earlier example: John has a reason to drive his car to the station because by going to the station he will make James happy. This reason statement includes references to each of the elements of the complete reason, i.e., the agent (x_1): John, the action (a_1): driving his car to the station, and a valuable outcome (o_{v1}): making James happy. However it is also true that John has a reason to drive the car that is parked in his garage to the station because by going to the station he will make James happy. This reason statement also includes references to each of the elements of the complete reason, i.e., the agent (x_2): John, the action (a_2): driving the car that is parked in his garage to the station, and a valuable outcome (o_{v2}): making James happy.

Since it is John's car that is parked in his garage, the complete reason $R''(a_1), o_{v1}, x_1$ is identical with the complete reason $R''(a_2), o_{v2}, x_2$. Yet if a_1 is not identical to a_2 , then CRIR would not hold. So identity between elements of complete reason statements is not a necessary condition for two reason statements to refer to the same reason. This will be referred to as the false negative challenge, because CRIR yields false negative result that the complete reason $R''(a_1), o_{v1}, x_1$ is *not* identical with the complete reason to by $R''(a_2), o_{v2}, x_2$.

Yet this is surprising, because CRIR appears to be trivially true. Complete reasons, on this account, are constituted by the elements standing in a reason relation. When the same elements stand in the reason relation they constitute a reason, i.e., one and the same reason. The issue here is that CRIR turns on the identity relation between elements of complete reasons. This helps to locate the question. The question is how to establish the identity between elements of complete reasons? Or, conversely, the question is how to distinguish between two different actions and/ or two different outcomes?

Both of the previous examples turned on facts not explicitly referenced in R'' . This may suggest that R'' is in some way incomplete. If these facts had been included in R'' CRIR would have held. However, there is a difficulty with this thought. As was discussed earlier, there may be no end to specification. In these case there would be no complete articulation of R'' , which would not possibly be subject to further augmentation through the addition of another fact. This is a poignant challenge, for if there is no complete articulation of R'' , then

there may not be a systematic basis for establishing the identity or difference of reasons. Identity statements about reasons, even complete reasons, would be dispositive.¹¹⁶

Rather than propagating the facts that are included in the reason relation, there is another possibility. It is worthwhile considering the ontology of the elements of complete reasons. What type of things are the elements of complete reasons such that they can be described in such different ways?

Indeed the root of the problem seems to be that the non-evaluative aspects of the action and the outcome admit to different levels of description in a way that is similar to evaluative facts. For present purposes this will be referred to as level flexibility. Yet, facts themselves do not stand in a specification relation to one another. Facts do not share level flexibility. This suggests it may not be fruitful to think of actions and events of having the ontological character of facts. What is the ontology of the action and the outcome?

¹¹⁶ Raz suggests that this is the case for explanations of reason statements. "As you see I am using 'reason' to refer to any fact which figures (nonredundantly) in an explanation, and not merely to the totality of facts all of which figure (non-redundantly) in an explanation. It is tempting to call the totality of all the facts which figure non-redundantly in an explanation a complete reason. I may occasionally use the term in order to avoid complex formulations. But if taken literally it implies more than is warranted: it implies that there is at least one comprehensive way of individuating facts, such that relative to any such scheme of individuation an object of explanation, it is either true or not, regarding each fact, that it belongs to the explanation of that object. There is reason to doubt that the explanation relation is such that it is ever true that regarding any object of explanation there is a set of explanatory facts such that it explains that object, and that adding any other fact to it is redundant so far as that explanation goes. It seems that our ways of individuating facts and the notion of explanation are such that any explanation can always be nonredundantly amplified, clarified and expanded." "Reasons: explanatory and normative", Raz, pg 3. The difference between Raz's point and the issue here, is that the current focus is on reasons, and the identity statements that hold between them, rather than on explanations of reasons statements. Even if explanations of reason statements may always be subject to augmentation, one may wonder whether reasons themselves are. Indeed, the following discussion will suggest an approach for addressing the fact that the elements of complete reasons can be referenced with so many different statements of fact.

The common description of reasons as facts is often meant to offer a contrast with views of reasons as statements or beliefs, i.e., with reasons as “what is the case”¹¹⁷ rather than as “what they (agents) believe to be the case”¹¹⁸ or what agents say. Facts in this sense are to be understood broadly to include: “...the occurrence of events, processes, performances and activities.”¹¹⁹ The ontology of the elements of complete reason then is an open question.

In keeping with the spirit of the current work, the following attempt to arrive at a plausible answer to this question will be offered in full recognition of the possibility that there may be more than one adequate way of accounting for the ontology of the elements of a complete reason. Indeed exploring the ontology of the elements of complete reasons raises questions about the delineation of elements of complete reasons offered so far. Is it appropriate and useful to distinguish between the agent, the action, and the valuable outcome? Why should one think that this division carves things at the joints?

Reflecting on the ontological characteristics of the elements of complete reasons will go some way to allaying these concerns. Practical reasons are action guiding. They are, frequently, reasons for an agent to act with intention to bring about an outcome.¹²⁰ Indeed, the elements discussed can be distinguished by their temporal markings. While the agent persists through

¹¹⁷ Raz, pg 17.

¹¹⁸ Raz, pg 17. Parenthetical added

¹¹⁹ Raz, pg 18

¹²⁰ At least this is the case for the subset of reasons of interest here.

time, the action and outcome happen at a time. While the agent may persist throughout the outcome, with the onset of the outcome the action (begins to) expires. As a result, the agent can be distinguished as ontologically distinct from the action and outcome. On the other hand, the action and outcome can be distinguished from one another by temporal order and the relevance of intentionality.¹²¹ Even if this suggests that the elements of complete reasons have different temporal characters, it leaves open the question of the ontology itself. However, it does suggest that a viable ontology for the present account of complete reasons respect these differences.

A plausible ontology should serve Raz' original motivation of speaking of reasons as facts, i.e., that which is the case, respect the different ontological character just discussed and be consistent with the explication of pragmatic consideration offered. Given the most recent discussion, the ontology of events Davidson relied on in "Paradoxes of Irrationality"¹²² is one possibility.

Events for Davidson are particulars, i.e., that which is the case. As particulars, events admit to descriptions at a multitude of levels, as an infinite number of facts are true of any event. This fits well with the fact that elements of a complete reason can be referenced at many different levels of specificity. Particulars share the characteristic level flexibility with the elements of a complete reason.

¹²¹ With the exception, of course, of cases discussed earlier in which the action is valuable in and of itself.

¹²² Davidson, Donald "Paradoxes of Irrationality", 1982, in *Problems of Rationality*, Oxford University Press.

The remainder of this section will explore the potential of this approach. There is a real possibility that it may ultimately prove too limiting to suggest that one can provide an account of reasons based on an ontology of events. If so, nothing above should turn on this. Indeed there is a problem with the suggested ontology. Elements of complete reasons seem to have a second dimension of flexibility that Davidson's particulars do not share.

Earlier it was mentioned that the particulars Davidson mentioned fit well with the fact that elements of a complete reason can be referenced at many different levels of specificity. Still, the action and outcomes elements of complete reasons have a different dimension of flexibility that particulars do not share. For example, consider the case of John's driving to the station. While the particular in which John drives to the station wearing a blue hat is different than the particular in which he drives to the station wearing a red hat, both are elements of John's reason. That is to say that it is not the case that John has a reason to drive to the station wearing a red hat and a further reason to drive to the station wearing a blue hat. John has one and the same reason to drive to the station whether he is wearing a red or a blue hat. Indeed there are possibly an infinite number of such variations. And the action that John has a reason to undertake is possibly consistent with an infinite number of different particulars. Thus it seems particulars are a poor candidate for the ontology of the action and outcomes of a complete reason. Does this mean that the ontology of events should be discarded out of hand? If not, how can it accommodate this

second dimension of flexibility, which for present purposes will be referred to as breadth flexibility, of actions and outcomes?

Action types and event types¹²³ share breadth flexibility while maintaining the level flexibility of particulars. For example, the particular in which John drives to the station wearing a blue hat and the particular in which John wears a red hat are tokens of the event type John drives to the station. As an event type, John's driving to the station is also consistent with an infinite number of different particulars. E.g., John's driving his car and John's driving the car that is parked in his garage are also tokens of this type. So event types share the characteristic level and breadth flexibility of elements of a complete reason.

This raises the question of how to establish identity between event types, but a detailed discussion of identity relations amongst event types would take things far afield from a discussion of reasons. One possibility will be offered later, but it is acknowledged that this approach relies on the claim that such identity conditions between event types can be established. If they cannot, this approach will have to be adjusted.

¹²³ From here on action types and event types will be referenced as event types. This is not intended to signal that the events are not actions in Davidson's sense. Presumably most if not all will be. But rather it will be helpful to distinguish action types from the action elements of complete reasons. Further, since actions, for Davidson, are events that are intentional under some description it is not inaccurate to describe them in this way, even if it does so obscures the intentional character of the action.

There is a related question which pertains more directly to reasons which are the current focus of discussion. How are the event types defined?

The definition of event types can also occur at different levels. Consider the following example. The agent has reason to go to the marionette theater since doing so would be valuable. This is ambiguous between two different reasons that the agent has. The agent has reason to go to the theater to perform as a puppeteer, or manipulator as they are called, since it would provide the agent a valuable opportunity to practice the craft he has long studied. The agent has reason to go to the theater to as an audience member since it would be valuable to engage with the artful production. The first reason statement is ambiguous between the two subsequent reasons, which in turn are exclusive of one another. I.e., the agent can either practice his craft or can engage with the artful production, but he cannot do both. This is not an issue that stems from the current analysis of reasons. It is a natural consequence of the specification relation that holds amongst values. Recall the specification relation: "If E , F and G are variable for properties, then F specifies G iff:

- 1) necessarily all F 's are G 's, but
- 2) possibly some G 's are not F 's, and
- 3) if a G is not and F , then necessarily there is some E such that necessarily all E 's are G 's, but possibly some G 's are not E 's."¹²⁴

¹²⁴ Heuer, pg 20.

F and E are complete subsets of one G, but some F's may be exclusive of E's. In this case a statement of G would be ambiguous between F's and E's (and possibly others).

While this issue does not stem from the current analysis of reasons, it is an issue that the analysis should accommodate; and it raises a question about the level at which event types are defined. For, if the action, for example, is defined at the too high a level it may also be ambiguous. An extension of this example makes the problem clearer. The agent has reason to go to the theater since it would provide the agent a valuable opportunity to practice the craft he has long studied. As a reason statement this may be unproblematic, but at the level of the complete reason, it is. If the action is defined as the event type, going to the theater, the reason relation, R'' , will not hold. In this case R'' would take the following arguments: a: going to the theater, o_v : practice the craft he has long studied, x: the agent. Using the plain language description of R'' this can be articulated as: the agent has a reason to go to the theater if doing so would bring about or increase the likelihood of the agent's practicing the craft he long studied. But there are tokens of the type "going to the theater" in which the agent goes to the theater as an audience member. And, as per the example, in these instances the agent does not have an opportunity to practice the craft he long studied, so the reason relation would not hold. Is this a problem for the current analysis of complete reasons? Are event types too flexible?

On the contrary, this illustrates one of the implications of the conditions of the reason relation. Recall the conditions of the reason relation:

- Conditions:

- The reason is a reason for the agent, x , to take the action, a .
- The action must bring about, or at least increase the likelihood of the outcome, o .
- The valuable property, v , is realized in the outcome, o .

The second and third conditions impose a requirement among elements of a complete reason. If the event type of the action is defined at higher a level than other elements in the complete reason, it may include tokens that will not bring about or increase the likelihood of the valuable outcome. And if the event type of the outcome is defined at higher level than other elements in the complete reason, it may include tokens that will not realize the valuable property. Note this requirement is asymmetrical. In other words, if the event types of the action and outcome are defined at a lower level than other elements in the complete reason they will include only tokens that bring about or increase the likelihood of the valuable outcome, and tokens that realize the valuable property, but they will exclude others that do as well.

This is the result of the relation that holds between the relevant event types. It has already been acknowledged that event types can be defined at different levels. Consider two event types, one defined at a higher level and one defined at a lower level. For the event type defined at a lower level to be consistent with the event type defined at a higher level the tokens of the lower level event type must be tokens of the higher level event type. For if it did not, the lower level event type would include tokens that were not tokens of the higher level event type and so would be inconsistent with it. On the other hand, the lower level event type need not include as tokens all the tokens of the higher level

event type. If there are tokens of the higher level event type that are not also tokens of the lower level event type, there is presumably another lower level event type of which they are tokens. This is just the specification relation. So the specification relation holds amongst relevant event types.

This makes it relatively easy to enumerate a consistency condition that reflects the consistency condition imposed by the second condition of the reason relation. Specification Consistency: For the reason relation, R'' , to obtain the event types which comprise the action and outcome must be specified at a level that is at least as low as the specification of the evaluative property.

Articulating Specification Consistency, it is easy to see that event types are not too flexible. The current analysis can accommodate the fact that the specification relation holds amongst evaluative properties because the specification relation also holds between the relevant event types. The reason relation, R'' , simply requires that the evaluative element and event types of the action and outcome elements to be specified at consistent levels.

But what of the asymmetrical character of this requirement? If the event type of the action is described at a lower level than other elements in the complete reason then it will exclude tokens that bring about or increase the likelihood of the valuable outcome. Similarly if the event type of the outcome is described at a lower level than other elements in the complete reason then it will exclude tokens that realize the valuable property. Neither would be a full articulation of the complete reason.

To this point, the assumed standard for completeness that anchored discussion of complete reasons and the reason relation was that the reason relation should include all the arguments required for reasons to be the reasons that they are. The standard of completeness was, so to speak, analytic. In contrast there is another standard of completeness which would be extensional. A complete reason relation would be extensionally complete if it applies to all cases in which the reason would apply.

Whereas Specification Consistency is required by analytic completeness, extensional completeness can anchor a different requirement. For the reason relation to be extensionally complete the event type must be defined at a high enough level that it includes all the tokens that will bring about or increase the likelihood of the valuable outcome. And the event type of the outcome must be defined at a high enough level that it includes all the tokens that realize the valuable property. Since it was already shown that if the event types of the action and outcome are defined at a lower level than other elements in the complete reason they will exclude tokens that bring about or increase the likelihood of the valuable outcome, and tokens that realize the valuable property, it is relatively easy to enumerate a condition that that reflects the Extensional Completeness: For the reason relation, R , to be extensionally complete the event types which comprise the action and outcome must be specified at a level that is at least as high as the specification of the evaluative property.

The conditions of R'' can be restated as follows:

- Conditions:
 - The reason is a reason for the agent, x , to take the action, a .
 - The action must bring about, or at least increase the likelihood of the outcome, o .
 - The valuable property, v , is realized in the outcome, o .
- Requirements:
 - Specification Consistency: For the reason relation, R'' , to obtain the event types which comprise the action and outcome must be specified at a level that is at least as low as the specification of the evaluative property.
 - Extensional Completeness: For the reason relation, R'' , to be extensionally complete the event types which comprise the action and outcome must be specified at a level that is at least as high as the specification of the evaluative property.

With an understanding of the elements of complete reasons as event types, Specification Consistency and Extensional Completeness address both the false negative and false positive challenges to CRIR.

Recall the false negative problem arose when the action element of $R''(a_1), o_{v1}, x_1$ was specified at a different level than the action element of $R''(a_2), o_{v2}, x_2$, thus CRIR yielded the false negative result that $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ were different reasons.

Assume that $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ satisfy Specification Consistency and Extensional Completeness. As a result, the elements a_1, o_{v1}, x_1 are specified at the 'same' level of specification as one another; and the elements a_2, o_{v2}, x_2 are defined at the 'same' level of specification as one another.

Further if the specification of the evaluative element of $R''(a_1), o_{v1}, x_1$ is specified at the same level as the evaluative element of $R''(a_2), o_{v2}, x_2$, then each of the other elements of the $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ are also specified at the same level of specification of their counterparts. Further, note that if two consistent even types, i.e., event types that stand in a specification relation to one another, are specified at the same level, then they share the same tokens. For if they do not, then they are either not specified at the same level; or they are not consistent and represent different event types. Further, two event types that share all the same tokens just are the same event type. Together this leads to the following conditional claim:

- if $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ independently satisfy Specification Consistency and Extensional Completeness
- and if the evaluative elements of $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ are specified at the same level,
- then the complete reason $R''(a_1), o_{v1}, x_1$ is identical with the complete reason $R''(a_2), o_{v2}, x_2$ if and only if $a_1 \equiv a_2, o_{v1} \equiv o_{v2},$ and $x_1 \equiv x_2$.

Since CRIR just is the consequent, CRIR holds if the conditions of the antecedent obtain. When the Specification Consistency and Extensional Completeness requirements are met, CRIR does not face the false negative challenge.

Specification Consistency and Extensional Completeness also show why the false positive challenge is not a challenge that plagues CRIR, but is a matter of interpretation of reason statements. Recall the false positive challenge emerged when two different reason statements were rendered at too high a level of specification. The reason statement “Mary has reason to go to the symphony if Mary’s attending the symphony would bring about or increase the likelihood of Mary’s engaging with the artistry of the music” was regimented in the following way: Mary has reason to go to the symphony, if attending the symphony would be valuable. This reason statement includes references to each of the elements of the complete reason, i.e., the agent (x_1): Mary, the action (a_1): going to the symphony, and a valuable outcome (o_{v1}): attending the symphony would be valuable. However, the different reason statement, which apparently referred to a different reason, “Mary has reason to go to the symphony if Mary’s attending the symphony would bring about or increase the likelihood of Mary’s engaging with the artistry of the conductor” was regimented in the following way: Mary has reason to go to the symphony, if attending the symphony would be valuable. This reason statement includes references to each of the elements of the complete reason, i.e., the agent (x_2): Mary, the action (a_2): going to the symphony, and a valuable outcome (o_{v2}): attending the symphony would be valuable. It is the case that: $a_1 \equiv a_2$, $o_{v1} \equiv o_{v2}$, and $x_1 \equiv x_2$. The complete reason $R''(a_1, o_{v1}, x_1)$ is identical with the complete reason $R''(a_2, o_{v2}, x_2)$, yet the reason statements refer to two different reasons. How are we to understand this?

The discussion of Specification Consistency and Extensional Completeness sheds light on the problem. The elements of $R''(a_1, o_{v1}, x_1)$ and $R''(a_2, o_{v2}, x_2)$,

take as tokens all and only the tokens that realize the valuable property of attending the symphony. This includes the value of engaging with the artistry of the music and the value of engaging with the artistry of the conductor. That $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ are ambiguous between two reason statements is not a challenge for CRIR. CRIR holds. $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ are identical reasons. It is just that they are reasons for more than one action that bring about or increase the likelihood of more than one valuable outcome. This is a challenge of interpreting reasons statements, not a challenge at the level of complete reasons. For CRIR to be fruitful, it will be helpful to interpret reason statements at the appropriate level of specification. As was mentioned earlier, there is likely no correct level of specification of a reason. The level of specification appropriate will be determined by the reasons one is comparing.

It is not the case that identity statements about reasons, even complete reasons, are dispositive. CRIR can establish the identity or difference of two complete reasons $R''(a_1), o_{v1}, x_1$ and $R''(a_2), o_{v2}, x_2$ that satisfy Specification Consistency and Extensional Completeness. This is based on understanding the ontological character of actions and outcomes as event types. And fruitful application of CRIR will depend on the appropriate level of specification of the elements of complete reasons in question which can be determined by context and pragmatic considerations.

This second stage of this chapter sought to make discussions of the remarkably sophisticated basic normative concept reasons tractable by focusing on an analysis of reasons themselves. In particular, to complement the earlier discussion, this stage targeted three aspects of reasons discussions which can

be troublingly opaque, i.e., when is something a reason, what is the relationship between the elements of a reason and how does one establish the identity or difference of two reasons?

The discussion began with a candidate for the complete reason relation suggested by Raz, $R(\Phi)p,x$, but argued that it did not meet a pertinent desiderata for an analysis of reason statements because it obscures what is required for p to be a reason to Φ . Noting that the earlier discussion showed that a number of different facts may play a role in a given reason, a variation of Raz's formulation, $R'(\Phi)p,c,x$ was offered. Unfortunately $R'(\Phi)p,c,x$ still obscured what it is about the relations between Φ , p , c and x that make it the case that the reason relation obtains. So a further variation of Raz's formulism, $R''(a),o_v,x$, was offered which can be articulated as : The reason relation obtains if an agent's, x 's, taking an action, a , brings about or increases the likelihood of an outcome, o , which realizes or preserves some value, v . Unlike R and R' , R'' wears the logic of the reason relation on its sleeve and makes explicit what is required to be an argument of the reason relation. The reason relation and its attending conditions were initially crystallized as follows:

- Formal analysis: $R''(a),o_v,x$, where R'' is the reasons relation, a is an action, o_v a valuable outcome, and x is an agent.
- Conditions:
 - The reason is a reason for the agent, x , to take the action, a .
 - The action must bring about, or at least increase the likelihood of the outcome, o .

- The valuable property, v , is realized in the outcome, o .
- Plain language description: The reason relation obtains if the agent's, x 's, taking an action, a , brings about or increases the likelihood of a valuable outcome, o_v .

With the suggested view in hand, several challenges to it were explored. To start with it was acknowledged that the suggested view differed from the previous discussion both in form and in tense, i.e., the previous discussion spoke of reasons in the present tense, whereas the suggested view speaks of elements of reasons as future tensed. Upon closer examination, it was suggested that difference in tense may be merely superficial. For the previous description of reasons also included elements that were temporally elongated. The differences in form may reflect a difference in their purpose. Whereas the previous descriptions of reasons sought to clarify which element provides the reasons, R'' seeks to elaborate upon the relationship between the elements of a reason.

Still, it was acknowledged that the broadly speaking consequentialist representation of reasons by R'' presents three challenges. First, there are some actions which are valuable in themselves. Second, there are other actions in which the character of the action is relevant to the reason. And third, it is highly likely that deontological considerations also provide reasons.

After looking at each of these in turn, it was argued that the first two challenges do not present a problem for the suggested view. For by permitting the outcome, o , to take as an argument a description of the action which

incorporates the evaluative element, R" can accommodate cases in which an action is valuable in and of itself. And cases in which the character of the action are relevant to the reason simply require that this aspect of the action is incorporated into the specification of the action element of R". However, the third challenge shows a limitation of the suggested view. If there are deontological considerations which provide reasons, and there likely are, then the suggested view should be augmented to account for them. In its current incarnation the suggested view pertains to value based reasons.

Understanding the significance of these challenges and noting the fact that the suggested view meets a desiderata for an analysis of complete reasons, the discussion explored whether R" could be used to shed light on complications that arise from the pragmatic way reasons are often discussed. In particular, can R" be used to offer a principled basis to determine whether two reason statement refer to the same or different reasons?

In considering this possibility it was noted that two different fact statements may refer to one and the same reason and that a single fact statement can be used to refer to more than one reason. Since the reason relation was fruitful in explaining how this could happen, it was suggested that it may be similarly fruitful as a basis for identifying and distinguishing reasons. This suggestion took the shape of the Complete Reason Identity Relation (CRIR): the complete reason R" $(a_1), o_{v1}, x_1$ is identical with the complete reason R" $(a_2), o_{v2}, x_2$ if and only if $a_1 \equiv a_2$, $o_{v1} \equiv o_{v2}$, and $x_1 \equiv x_2$.

Unfortunately it was quickly recognized that CRIR appeared to face both a false negative and a false positive challenge. It was noted that both challenges turned on facts not explicitly referenced in R". However, since there may be no end to specification it may always be possible to augment R" through the addition of another fact. In these case there would be no complete articulation of R" and as a result there may not be a systematic basis for establishing the identity or difference of reasons. This is a poignant challenge because identity statements about reasons, even complete reasons, would be dispositive.

Rather than seeking to address this challenge by propagating the facts that are included in the reason relation it was suggested that it may be worthwhile to reconsider the ontology of the elements of complete reasons.

The root of the problem was identified as the fact that the action and the outcome of the complete reason admit to different levels of description in a way that is similar to evaluative facts, which was referred to as level flexibility.

Since facts do not share level flexibility with actions and outcomes, it was suggested that they may not be a fruitful candidate for the ontology of elements of a complete reason. Indeed, it was noted, the original motivation for describing reasons as facts was not so much to ascribe an ontology to the elements of a complete reason but to distinguish them from beliefs or statements. An alternative ontology which respected this difference would be in keeping with the original motivation.

Inspired by Davidson's ontology of events as particulars, an ontology of event types was proposed for the elements of complete reasons. And it was shown that event types share the characteristic level and breadth flexibility of actions and outcomes. Further it was shown that the specification relation, that holds amongst evaluative properties, holds amongst consistent event types.

Recognizing this, it was easy to enumerate two requirements of the complete reason relation:

- Specification Consistency: For the reason relation, R'' , to obtain the event types which comprise the action and outcome must be specified at a level that is at least as low as the specification of the evaluative property.
- Extensional Completeness: For the reason relation, R'' , to be extensionally complete the event types which comprise the action and outcome must be specified at a level that is at least as high as the specification of the evaluative property.

Based on an understanding of the ontological character of the actions and outcomes as event types, it was shown that CRIR can establish the identity or difference of two complete reasons $R''(a_1), o_{v_1}, x_1$ and $R''(a_2), o_{v_2}, x_2$ that satisfy Specification Consistency and Extensional Completeness. It is not the case that identity statements about reasons are dispositive. R'' can offer a principled basis to determine whether two reason statements refer to the same or different reasons.

--

This chapter sought to contribute to the larger goal of enumerating a conception of practical reasoning consistent with Preference Theory by focusing on the concept of reasons. For normative reason, on this conception, is the basic normative concept.

While there is much worthwhile discussing about normative reasons, this chapter stayed narrowly focused on three points that will be pivotal to the subsequent chapters.

Normative reasons, on this conception, are value based. While non-evaluative facts may be reasons, it is evaluative facts and value itself that provides reasons. Desires, preferences and even beliefs do not. While it was acknowledged clearly that this puts the current conception at odds with those often associated with Preference Theory, it was argued that this is not as large a detriment as it may at first appear.

Further, this chapter laid out a straightforward formalism for complete reasons based on an ontology of event types. While other formalisms can be provided, and other ontologies may suffice, the suggested view clearly establishes what is required for something to be a reason and articulated a relationship between elements of a complete reason.

The suggested complete reason relation also provided a means for establishing the identity or difference of reasons.

Cumulatively these points provide the means for addressing a number of questions not yet posed in this chapter. For example, it provides a means for understanding at least three different ways that reasons can compete based on kind of value, extent of value and likelihood of realizing the valuable outcome. The significance of this will become clearer in subsequent chapters.

If this chapter contributed to the larger goal of this dissertation, it ultimately ended with an outstanding promissory note of explaining the relationship between reasons and preferences. Payment of this note is forthcoming, but will first require a discussion of rationality, the topic of the next chapter.

Chapter 3: Rationality

If reason¹²⁵ is the basic normative concept, rationality is the coin of the realm of preference theories. Preferences, after all, are said to be rational or irrational. But what is the relationship between reasons and rationality? What makes certain preferences rational and justifies the conditions of rationality? This discussion will seek to offer an account of rationality that can serve as the basis for answering these questions.

Rationality is central to many areas of study such as action theory, ethics and economics. The fact that each of these fields discusses different characteristics of rationality may suggest to some that these fields employ different concepts of rationality. In keeping with the spirit of the current work, this project goes in a different direction. This chapter will seek to articulate a conception of rationality that can serve to anchor Preference Theories yet draws from developments in action theory and ethics.

To begin, it will be helpful to be more specific about the notion of rationality this chapter is concerned with. For example, humans are described as rational, but dogs and cats (on most accounts) are not. On different accounts this rationality is attributed to a rational capacity or the fact that people are rational creatures. Though a person acts irrationally or has an irrational belief, she may nonetheless remain a rational creature or retain her rational capacity. This paper will be concerned with the rationality of people's actions, not with the kind

¹²⁵ NB: 'reason' here refers to reasons, not to the faculty of reason.

of creatures they are nor their rational capacity, except in so far as the two are connected.¹²⁶

The concept of rationality that will be dealt with here has been called subjective rationality to contrast it with objective rationality. This label can be misleading and should not be read to mean that the question of whether an agent's action is rational is not an objective fact. Instead, it points to the fact that the rationality of an agent's action is determined by the contents of the agent's mind, as it were, and not whether the agent arrives at the correct judgments based on the reasons that he has, for example.

Naturally this raises a question of why this is the focus of the current chapter. First, as was mentioned in Chapter 1: Introduction, the subject of this dissertation is a conception of practical reasoning, which in turn is comprised of conceptions of subsidiary concepts that play distinct but complementary roles. Second, the conception of reasons enumerated in the Chapter 2: Reasons addresses many of the motivating factors for an objective view of rationality. In other words, in many instances, questions of objective rationality can be cashed out in terms of reasons. As a result, using an objective understanding of rationality here would be duplicative and confusing.

Further, the objective account of reasons leaves much territory to be explained. For example, consider an agent who would be made better off by choosing A over B, and who has all things considered reason to choose A over B because of this. Yet the agent arrives at the judgment that choosing B over A would

¹²⁶ While the (ir)rationality of action will be the ostensive focus of our attention, parallel arguments can be offered for the irrationality of beliefs.

make him better off, and concludes he has most reason to choose B over A for this reason. The agent's reasoning is faulty.

On an objective account of rationality, some may claim his conclusion is irrational, for choosing B would not make him better off. This fault can also be explained by noting that the agent's judgment is not in accord with the balance of reasons.

Further, imagine that despite his judgment to the contrary the agent chooses A over B. Despite making the right choice, the choice in line with the balance of his reason, the agent's choice is also faulty in a way. This is not a fault of being objectively irrational. After all, the agent chose the option he had most reason to choose. This fault can be described as a fault of subjective rationality which is distinct from choosing in accord with the balance of reasons and reflects a fault in how the agent proceeded from judgment to choice.

Employing a subjective account of rationality in concert with an objective view of reasons offers the benefit of this explanatory breadth and clarity.

This is not intended as a refutation of objective views of rationality. And this is not an argument about vocabulary. If others seek to use 'rationality' to refer to objective rationality, nothing herein argues against the ability to do so. It is only an explanation of the view of interest here and an indication of a reason that recommends this choice.

It should also be noted that the choice between objective and subjective views of rationality are likely not exhaustive. Parfit, in On What Matters¹²⁷, appears to employ a compound view of rationality according to which an agent can be perfectly rational, i.e., subjectively and objectively rational, or partially rational, i.e., either subjectively or objectively rational. While this may be a plausible way of dividing the terrain, it seems to run a risk of difficult ambiguity.

As a result the present focus will remain on subjective rationality. It is this understanding of rationality that specifies requirements that are violated when an agent knowingly holds contradictory beliefs, fails to take the means he deems necessary to an end he intends, or is akratic. It is the understanding of irrationality exemplified in the following example.

To commit the murder he intends, the murderer believes it is necessary to buy an axe. By not forming the intention to buy the axe, the murderer is at fault.¹²⁸ This is true even if the murderer has no reason to buy the axe, because for instance, he has no reason to murder the victim. This cannot be the fault of failing to comply with the reasons he has. Rather the murderer is guilty of being irrational.

In addition to indicating the concept of rationality this project is concerned with, this example also illustrates one of the more informative strategies to pursue in exploring rationality. That is to focus on failures of rationality and draw out from these the distinctive features of rationality. This is a well established approach and one the current effort will adopt.

¹²⁷ Parfit, On What Matters, Oxford University Press, 2011.

¹²⁸ This is a variation of an example Raz suggests in "The Myth of Instrumental Rationality", *Journal of Ethics & Social Philosophy*, Vol. 1, No. 1, April 2005.

This discussion will proceed in three stages. The first will provide background by reviewing three very different approaches to irrationality which each make seminal contributions. The second will enumerate the suggested account. And the third will explore whether the suggested view can account for cases of irrationality.

Stage 1: Background

In setting out to review these treatments of irrationality it will be useful to reflect on the criteria an adequate account should meet for present purposes. This can serve as something of a guiding star for subsequent discussions.

Over time, as the focus on irrationality has progressed, so too have the goals for offering an account of irrationality. For example, a chief concern of Davidson's "Actions, Reasons and Causes"¹²⁹ was to explain how reasons could be causes, or how the mental could be causally efficacious. Discussion of irrationality in this context sought to explain the riddle posed by akratic choices: how an agent could intentionally act, i.e., act for a reason, contrary to his reasons all things considered. As a result, an account that rendered the possibility of irrational action intelligible, as Davidson's account did, was an adequate account of irrationality.

This account will have to go further. The goal is to offer an account of rationality that is distinct from, but complementary to, reasons, which can serve to anchor Preference Theories by serving as basis to justify the completeness

¹²⁹ Davidson, "Actions, Reasons and Causes", *The Journal of Philosophy*, Vol. 60, No. 23, American Philosophical Association, Eastern Division, Sixtieth Annual Meeting (Nov. 7, 1963), pp. 685-700.

and transitivity conditions. As a result, it should offer an account of the kind of failures that constitutes irrationality which are common to all and only to cases of irrationality. Further, since rationality is distinct from reasons, which is the basic normative concept, this account should explain why cases of irrationality often carry the hallmark of normativity.

An account of irrationality that explained a failure that was common to all and only cases of irrationality, yet was distinct from the failure to comply with reasons, and did so in a way that described the unique normative force of irrationality would be sufficient in describing how rationality is complementary to reasons.

With this background, two requirements for an account of irrationality can be offered, which will collectively be referred to as the Riddle of Irrationality:

1. The Descriptive Requirement: It will offer a descriptive account of the failure of irrationality that is common to all and only cases of irrationality.
2. The Normative Requirement: This common failure needs to be of a special kind that can explain the normative force of irrationality.

These requirements offer criteria to evaluate accounts of irrationality. The literature on the subject is too extensive to deal with exhaustively. A review of how the works of Davidson, Scanlon and Broome fare in meeting these requirements will help provide an understanding of what has yet to be accomplished. Each of these works, it will be argued, fails to meet the combination of the Descriptive and Normative requirements. That they do so reflects the force of these requirements and further recommends their

significance. How they do so, points to benefits of the account that will be suggested subsequently.

This stage will proceed in four sections, looking at the work of Davidson, Scanlon and Broome respectively before returning for a brief summary on how each fared in answering the riddle of irrationality.

Davidson

In “Actions, Reasons and Causes”¹³⁰ Davidson produced seminal work that offered a descriptive account of irrationality that is common to cases of practical and theoretical rationality, specifically to cases of weakness of the will (or akrasia) and weakness of the warrant. However, in order to meet the Normative Requirement, he had to elaborate on his descriptive account by positing the existence of an unlikely second order principle, i.e., the Principle of Continence, that agents were supposed to have adopted. The violation of the Principle of Continence was supposed to anchor the normativity of irrationality.

In addition to the fact that this principle was descriptively inadequate, in that it produced problematic results¹³¹, it did not have the resources to explain the normativity of irrationality. Specifically, it could not answer why violations of this particular principle carried the normative force of irrationality. To understand this, it will be necessary to look at Davidson’s work in more detail.

¹³⁰ Davidson, “Actions, Reasons and Causes”, *The Journal of Philosophy*, Vol. 60, No. 23, American Philosophical Association, Eastern Division, Sixtieth Annual Meeting (Nov. 7, 1963), pp. 685-700.

¹³¹ For example, on Davidson’s account an agent who did not happen to accept the principle could not be guilty of irrationality

In deliberating about whether to S, an agent considers her reasons for S-ing. She may believe she has reasons for and against S, though the conditional form of the agent's attitudes, Davidson suggests, ensures that even in such cases the agent will not thereby be entertaining a contradiction. Following Davidson, we can call these attitudes prima facie judgments (PF-J's), and they can be represented as follows: pf(evaluative judgment; evidential grounds) or as pf(action type; reasons).

An agent may simultaneously arrive at conflicting PF-J's such as pf(s, r_1) and pf(not s, r_2). In such cases, the agent resolves the conflicts by arriving at an all things considered judgment (ATC-J), which is a single conditional attitude based on all the relevant reasons, i.e., pf(s, r_1 and r_2).

In contrast, intentions, the attitudes on which an agent acts, are unconditional in form. Such intentions, or sans-phrase judgments (SP-J's), can be represented as follows: sp(s).¹³²

The question remains, how does an agent arrive at an SP-J from her PF-J's? Davidson offers an answer in the form of the Principle of Continence, which stipulates that an agent arrives at an SP-J based on the reasons he considered to arrive at his ATC-J.

¹³² There is a further contrast between PF-J's and SP-J's according to Davidson. Whereas PF-J's concern action types, SP-J's concern specific actions. Bratman suggested that this difference is a result of the emphasis Davidson places on present directed intentions, and creates difficulties for Davidson in dealing with future directed intentions. This difficulty can be overcome by taking SP-J's to range over action types as well. This position will not be argued for here, but because of this possibility nothing will be made of this difference between PF-J's and SP-J's on Davidson's account.

This discussion provides the basis for Davidson's account of akrasia.

Concisely put, in cases of akrasia an agent arrives at an SP-J from a subset of the reasons he considers in arriving at his ATC-J, which recommends an action different from the one recommended by his ATC-J. I.e., an agent arrives at an SP-J to S, to smoke for instance, based on r_2 though his ATC-J, recommends not smoking, based on r_1 and r_2 .

Davidson's descriptive account made a significant contribution to understanding akrasia. It shows how an agent can act intentionally, i.e., for a reason, though irrationally, i.e., against his better judgment. Further, in showing that his account could be applied to cases of weakness of the warrant as well as cases of weakness of the will, or akrasia, Davidson shows that his descriptive account picks out a failure of rationality that is common to two canonical cases of irrationality.

However, in order to meet the Normative Requirement, that is to show that the failure he points to can explain the normativity of rationality/irrationality, Davidson needs to say more, and for this he turned to the Principle of Contenance.

Davidson explains that it is in virtue of having this principle, that an agent's action is irrational. "If the agent does not have the principle that he ought to act on what he holds to be best, everything considered, then though his action may be irrational from *our* point of view, it need not be irrational from his point of view – at least not in a way that poses a problem for explanation."¹³³¹³⁴

¹³³ Davidson, Problems of Rationality, "Paradoxes of Irrationality" pg 177

With the riddle of irrationality in mind, Davidson can be read as saying that the normative force of rationality stems from the agent's violation of the Principle of Continenence. Specifically, he claims that it lies in the agent's ignoring of the principle because of a reason against acting on the principle, rather than on a reason against the principle.

The problem with this is that this kind of failure, if it is a failure, is not unique to cases of irrationality. Consider the following example. An agent may adopt honesty as a principle, yet when faced with a situation in which it would be advantageous to lie, the agent may choose to do so. Though the agent violated his principle based on a reason against acting on the principle rather than on a reason against the principle the agent is not thereby irrational, even if he is less than virtuous.

Since the failure that Davidson points to is not unique to cases of irrationality, his account ultimately fails to meet the Descriptive Requirement even though his account makes a seminal contribution to the understanding of how akratic action is possible.

Scanlon

For present purposes Scanlon's account of irrationality makes a very different contribution, as he appears to be the first in the literature to make an interesting suggestion regarding the normative force of rationality and irrationality.

¹³⁴ In characterizing the agent's action as possibly irrational from our point of view, though it is not from his, Davidson seems to slip into a wider conception of irrationality than will be employed later in this paper.

Unfortunately, as will be explored in the following discussion, it is unlikely that Scanlon's account will meet the Descriptive and Normative Requirements.

In "Structural Irrationality"¹³⁵ Scanlon posited that commitment is a constitutive part of intending¹³⁶. That is to say, in deciding to do A at T, or intending to do A at T, one commits oneself to take one's doing of A at T into consideration in subsequent deliberation. If one intends to do A at T, then, according to Scanlon it is irrational not to take the fact that B would contribute to doing A at T as a reason for doing B, since one has committed to do so. And, it is irrational to form the intention to A at T if one judges that one does not have sufficient or conclusive reason to A at T, since doing so would commit one to taking the fact that B contributes to doing A at T as a reason to do B, though one does not judge it to be such a reason. In both cases, the agent has committed himself to take a fact as a reason, though he does not judge that it is one. This then is how Scanlon caches out two locations of irrationality in terms of commitment.

The difficulty with Scanlon's account for present purposes is that it does not meet the Descriptive Requirement. For instance, arguably there are present-directed intentions, i.e., intentions to achieve an end that do not require an agent to take a further action as a means to the end. And such intentions can also be irrational. The constitutive role of commitment in intention Scanlon describes cannot account for this failing of irrationality¹³⁷. As a result, Scanlon's

¹³⁵ Scanlon, "Structural Irrationality" in Common Minds: themes from the philosophy of Philip Pettit, Oxford : Clarendon Press ; New York : Oxford University Press, 2007, pg 84–103.

¹³⁶ This is a feature that Scanlon shares with Bratman. Bratman, M., Intention, plans, and practical reason, Cambridge, Mass. : Harvard University Press, 1987.

¹³⁷ This is not to suggest that this is the only challenge that faces Scanlon's account. There are many others, but this is sufficient to note for current purposes.

account, does not describe a failure of rationality that is common to all cases of irrationality, and therefore does not meet the Descriptive Requirement.

Despite the limitations of Scanlon's descriptive account, he appears to be the first in the literature to raise an interesting explanation of the force of rationality¹³⁸.

Scanlon suggests that the normative force of rationality/irrationality stems from the tension between the agent's judgments¹³⁹ and what one takes to be a reason. For instance, in the first location of irrationality that Scanlon points to, it is irrational not to take the fact that B contributes to A as a reason to B, if one intends to A, because one has committed oneself to do so. The irrationality arises between one's intention and the reasons one takes them to be. The same issue can be pointed to in the second location of irrationality Scanlon points to. This is an account of irrationality based on the agent's reasoning and is quite independent from the question of what reasons there are.

For present purposes there is much promising in Scanlon's account, and the suggested account will share much in common with it, though it will suggest the source of the failure which constitutes irrationality is almost the reverse of the one Scanlon points to.

Broome

¹³⁸ Strictly speaking, Scanlon offers two different accounts of the normativity of rationality/irrationality, but for our purposes one can safely be ignored.

¹³⁹ For present purposes intentions will be understood as judgments. Though, if there is reason to question this categorization of intentions nothing should turn on this.

On the other hand, Broome's rational requirements provide an ingenuous answer to the Descriptive Requirement. Rather than pointing to an underlying failure that constitutes irrationality, Broome's rational requirements are a list of necessary conditions for rationality such that if an agent violates a rational requirement, then he is irrational. Here are four examples of Broome's rational requirements¹⁴⁰ that basically correspond to contradiction, modus ponens, instrumental reasoning, and an expansive version of akrasia:

- First requirement: Rationality requires of you that you do not both believe p and believe not-p.¹⁴¹
- Second requirement: Rationality requires of you that, if you believe p and you believe (if p then q), and if it matters to you whether q, then you believe q.¹⁴²
- Third requirement: Rationality requires of you that, if you intend to G, and if you believe your F-ing is a necessary means to your G-ing, and if you believe you will not F unless you intend to F, then you intend to F.¹⁴³
- Fourth requirement: Rationality requires of you that, if you believe you ought to F, and if you believe you will not F unless you intend to F, then you intend to F.¹⁴⁴

A complete enumeration of the rational requirements would presumably rule out every possible case of irrationality. Because such a complete list would specify a failing common to cases of irrationality, namely violating a rational requirement, it provides a plausible response to the Descriptive Requirement.

¹⁴⁰ Broome exercises care in the specific form of the rational requirements he presents, often for good reasons. However, this involves a level of granularity that is not required for current purposes. Where convenient minor changes have been made for consistency and readability.

¹⁴¹ Broome, B., "Reasoning", unpublished manuscript, 2005, pg 75.

¹⁴² Broome, B., "Reasoning", unpublished manuscript, 2005, pg 75.

¹⁴³ Broome, B., "Reasoning", unpublished manuscript, 2005, pg 77.

¹⁴⁴ Broome, B., "Reasoning", unpublished manuscript, 2005, pg 79.

Until relatively recently, Broome had a straightforward answer to the Normative Requirement. The Requires relation found in rational requirements was normative, he argued. The Requires relation was different from, but akin to the reason relation.

Recently, he seems to have backed away from this claim and begun to look for an argument to substantiate it. However, in order to offer a substantial response to the Normative Requirement Broome will have to answer the question: what do the rational requirements share in common such that they can explain the normativity of rationality?¹⁴⁵ As such an answer will need to point to a failing that is common in cases of irrationality, it will sacrifice one of the significant benefits of his answer to the Descriptive Requirement, namely, not pointing to a common failing of this kind.

Broome's response to the Normative Requirement comes at the expense of his ingenuous answer to the Descriptive Requirement. As such, it would seem that Broome's account in its current form can either meet the Descriptive or the Normative Requirement, but not both, and so does not provide a successful response to the Riddle of Irrationality. To be clear, this is not to say that Broome's rational requirements are inconsistent with a successful response to the Riddle of Irrationality¹⁴⁶.

¹⁴⁵ It is worthwhile questioning the how one determines how one should enumerate a list of the rational requirements. E.g., "How can we test whether some putative requirement is genuinely a requirement of rationality? It would be nice to have some general criterion to apply, or at least some general method. But I am sorry to say I do not have one. Several philosophers have argued that rational requirements must be somehow inherent in the nature of the mental states they are concerned with. I am sure they are right in some way." Broome, B., "Reasoning", unpublished manuscript, 2005, pg 71-72.

¹⁴⁶ Indeed, to the contrary, later it will be argued that Broome-style rational requirements are consistent with the account of rationality that will be suggested.

For example, as Ulrike Heuer and Christian Piller pointed out, there may be other approaches that provide the basis for augmenting Broome's account. Specifically they suggest a possibility that Broome mentions in "Reasoning"¹⁴⁷, namely that rational requirements must be somehow inherent in the nature of the mental states with which they are concerned. Could an account based on this insight successfully answer the Riddle of Irrationality?

This is certainly a possibility. While Broome himself has not developed this position, there are other accounts of this type that suggest this possibility may be worthy of consideration. Christine Korsgaard's account of instrumental reasoning is perhaps the most notable example.

While Korsgaard's account offers the basis for a distinctive response to the Normative Requirement – namely that the Instrumental Principle, an aspect of the Categorical Imperative, is constitutive of autonomy itself – it would need to be developed further to offer a response to the Descriptive Requirement.

The challenge here is to augment Korsgaard's account to apply to all and only cases of irrationality. For example, Korsgaard presents the Instrumental Principle as: "practical rationality requires us to take the means to our ends"¹⁴⁸. While this is in the general spirit of the third of Broome's rational requirements listed above, in its current form it is both too broad and too narrow. This is too broad for precisely the reasons that Broome's requirements are more specific. Namely that in many instances rationality may not require us to take the means

¹⁴⁷ Broome, B., "Reasoning", unpublished manuscript, 2005, pg 71-72.

¹⁴⁸ Korsgaard, Christine, Constitution of Agency, Oxford University Press, 2008, 5.

to our ends. For example, at present there may be no means we can take to accomplish our ends. In such a case, rationality would not require of us that which is not possible. It is exactly these types of specific considerations that Broome takes care to take into consideration when formulating his specific rational requirements.

Further, Korsgaard's Instrumental Principle is too narrow as the sole basis for rational requirements for it does not apply to other cases of irrationality such as cases of akrasia.

Neither of these are arguments against the possibility that Korsgaard's position could be developed to offer answers to the Descriptive and Normative Requirements. They only point out questions that would need to be answered in order to do so.

Setting this aside for future exploration, it is important to note that Broome's rational requirements have been fruitful in unearthing a number of features of rationality. Two are particularly relevant to the current discussion, specifically that unlike the reasons relation, the Requires relation is both broad and strict.

Consider a case of means ends reasoning. If an agent has a reason to intend an end, and a means is necessary to achieving that end, then the agent has reason to take the means. The Reason relation is narrow.

However, if an agent intends an end, and believes that a means is necessary to achieving the end, he is not required to take the means. Rather, what is

required of him is not to be in a state in which he intends the end, and believes the means necessary, but does not intend the means. He can avoid being in such a state by coming to intend the means, or by dropping the end, or by coming to believe that the means is not necessary. Any of the three will fulfil the Requires relation. The Requires relation is broad.¹⁴⁹

Further, while the agent has a reason to take the means, this is consistent with his not intending to take the means, because he may have more reason not to. The Reason relation is slack. In contrast, the Requires relation is strict. Though there are three different ways in which the agent can comply with the Requires relation in the current example, the strictness of the Requires relation demands that he do so in one of the three ways.

Even if Broome's account does not meet the Normative requirement, it looks like any account of rationality/irrationality that does will have to accommodate these features of the Requires relation.

Riddle of Irrationality

While none of these accounts answers the Riddle of Irrationality, they each shed light on important aspects of irrationality. In what follows the discussion will try to offer an account of irrationality that answers the Riddle of Irrationality by building on Davidson's descriptive account. The suggested account is consistent with the insight Scanlon expressed about the connection between the normativity of reasons and the normativity of rationality, and can serve as a

¹⁴⁹ It is important to note that there is significant debate about whether the rationality is broad. Unfortunately a full discussion of the topic goes beyond the scope of the current effort. The later discussion will point to the resources the suggested account offers to defend the broad scope position from the most prominent criticism.

basis for enumerating rational requirements that are similar in kind to those discussed by Broome.¹⁵⁰

Stage 2: Suggested Account – Rationality

This stage will proceed in three sections. The first section will concisely introduce the suggested account of rationality based on the notions of ‘recognizing something as a reason’ and ‘taking something to be a reason’. The second section will focus on the notion of ‘recognizing something as a reason’, addressing two key objections before exploring the unique character of recognition on which the account depends. The third section will focus on the notion of ‘taking something to be a reason’, addressing multiple forms of a key objection before reflecting on the nature of deliberation.

Suggested Account of Rationality

Rationality consists in being ruled by reason. To be ruled by reason, reason must be directive and efficacious.

Reason is directive when an agent recognizes that a consideration counts in favor of a proposition or an action; that is when *an agent recognizes a reason as a reason*.

Reason is efficacious when an agent takes a consideration to count in favor of a proposition or an action in deliberation; that is when *an agent takes a reason to be a reason*.

¹⁵⁰ Though the rational requirements that will be discussed differ than those discussed by Broome in that he posits that they are analytically prior to rationality whereas on the suggested account they are a consequence of rationality.

For an agent to be rational, to be ruled by reason, reason must be directive and efficacious. To be rational, the agent must recognize a reason as a reason, and take it to be the reason that he recognizes it to be.

To hold this view, more should be said about what it is to recognize something as a reason, and what it is to take something to be a reason.

Recognizing something as a reason

In exploring the possibility that rationality involves recognizing a reason as a reason, it will be helpful to address two likely objections before exploring the unique character of recognition.

The first objection imputes too much to the distinctive characteristics of recognition, and the second too little. Both will be looked at in turn.

The first objection stems from the veridical character of recognition¹⁵¹. In essence, the claim is that as a result of its veridical character, recognition is too demanding to play a meaningful role in an account of rationality.

The second objection has essentially the opposite force, claiming that recognition is not distinctive enough from belief to warrant a meaningful role in an account of rationality. I.e., any role that recognition can play can also be played by the concept of belief, and doing so has the added benefit of not needlessly propagating mental states. These objections will be addressed in turn.

¹⁵¹ This was first raised by Raz in conversation.

To clarify the force of the objection, it will be helpful to offer reflections on the veridical character of recognition. Roughly speaking, for an agent to successfully recognize something, the thing which is recognized must be the thing it is recognized to be. This point can be more easily grasped through example. Consider the case of two friends, Tom and Michael. In order for Tom to have recognized his friend in the crowd, it must be the case that Michael was in the crowd. Or, illustrating the other side of the veridical character of recognition, it would be inappropriate to claim that 'Tom recognized his friend in the crowd', if the friend in question, Michael, was not in the crowd. In such a situation, a more qualified version of the claim would be appropriate, such as 'Tom thought he recognized his friend in the crowd'. In this instance the act of recognition did not have the requisite success, and therefore this is not an example of recognition. This is reflective of the veridical character of recognition.

If, as a result of the veridical character of recognition, an agent can only recognize as a reason that which is a reason, then recognition may be a poor fit for an account of rationality. Recall the case of the axe murderer who could rationally intend to buy an axe, or irrationally fail to form this intention despite believing it necessary to achieve the end he intends. The axe murderer's rationality or irrationality is quite independent of the reasons he has (or in this case the reasons he does not have); for he has no reason to buy the axe. If it is the case that an agent must have the reason in question in order to recognize it as a reason, then recognition does not seem to play much of a role in rationality.

But this objection seems forceful because the objection itself makes too much of the veridical character of recognition. Consider a further variation of the example such that Michael was in the crowd. However, contrary to Tom's beliefs, Michael was not his friend. Michael had been merely pretending to be Tom's friend. In this case, it would not be inappropriate for Tom to say he 'recognized his friend in the crowd'; even if from a third person perspective it may be more appropriate to say 'Tom recognized someone he had mistakenly come to view as a friend'. Nonetheless there was an act of recognition.

The failure in this example lies somewhere else; not with the act of recognition, but with the background conditions upon which the recognition was predicated. The veridical character of recognition does not extend to the background conditions of recognition.

Plausibly there are other limits to the demands of the veridical character of recognition. For example Tom may have successfully recognized Michael even if he was mistaken to believe that there was a crowd. From a third person perspective it may be appropriate to say 'Tom recognized a friend in what he mistakenly believed to be a crowd'. There was a successful act of recognition, even though Tom had mistaken beliefs about the circumstances. The veridical character of recognition does not demand all such beliefs about the circumstances be true.

The veridical character is demanding, but limited. In order for Tom to have recognized Michael, it must be Michael that Tom recognized. But Tom's beliefs

about the background conditions and circumstances need not have been true, in order for Tom to have successfully recognized Michael.

This same point explains why the veridical character of recognition is not too demanding to play a role in rationality. However, when making the same point in respect to reasons, the limitations of reasons-language can create confusion. As a result it will be helpful to walk through the point step by step.

The previous chapter discussed the structure of complete reasons in some detail. Recall that it was suggested that complete reasons are comprised of elements, such as an agent, an action and a valuable outcome, which stand in a reason relation to one another. As a result, one can understand an example of a common reason statement such as: That I am thirsty gives me reason to drink that glass of water, because doing so would quench my thirst, in the following way:

- A agent: I (or me)
- An action: drinking that glass of water
- A valuable outcome: quenching my thirst.

To be a reason, each of these elements is necessary. Simply negating or changing one element or another should be enough to show this.¹⁵² For

¹⁵² It is important to remember two points from the discussion of the topic in the previous chapter. First there is more to be learned about the elements of reasons by considering which kinds of changes change the reason in question. I.e., some changes will just offer a different specification of the same reason, while others will offer a specification of a further reason, while others still will cease to make these elements a reason. Second, there are arguably many different ways to individuate the elements of a reason. Indeed individuation itself is a subject worth considerable attention at the level of beliefs, actions and events, not only at the level of reasons. Needless to say this involves considerable questions of ontology. The current effort does not presume to claim that the suggested approach represents the only way, or even a canonical way of individuating reasons and their elements. The claim here is merely that this is a fruitful and robust way for doing so. Further, any approach to individuating reasons and their

example: If drinking that glass of water would not quench my thirst, but, for the sake of the example, make me more thirsty, then my being thirsty would *not* give me reason to drink it. We can recognize the importance of each of the other elements by negating or excluding each of these in turn.¹⁵³

Further, the elements themselves are not sufficient to comprise a complete reason. The elements of a complete reason must also stand in the reason relation. To claim that ‘the fact that I am thirsty gives me reason to drink that glass of water in order to quench my thirst’, is a further claim than to claim: ‘I am thirsty, and drinking that glass of water will quench my thirst’. Present in the first statement, but not the second, is the claim that the reason relation obtains; i.e., that the elements of the reason stand in such a relation to one another that they can be arguments for the reason relation.

With this brief description of reasons, the discussion can return to the question of whether the veridical character of recognition is too demanding. Consider what would be involved in recognizing something as a reason. Here it will be helpful to make three points. To recognize a reason:

- it is not sufficient to recognize the reason relation, one must also have subsidiary propositional attitudes about the elements of the reason¹⁵⁴;

elements should offer an explanation of the unique normative force of a complete reason. That the suggested view meets this criteria, counts in its favor.

¹⁵³ This should not be taken to mean that in referring to a reason we need to refer to all of its elements. Many are simply supplied by context. So in the current example, if someone asked why I planned to drink the water, I might simply respond, that I was thirsty. My belief that the water would (partially) quench my thirst is understood from the context.

¹⁵⁴ If additional elaboration would be helpful, a variation of the earlier exercise can illustrate this point. Imagine the possibility of an agent reporting that ‘he had a reason’. When pressed to explain further, he merely insisted that he had a reason, but could not explain the reason with reference to any of its elements. For the present purposes assume the agent is not merely being coy – by relying on the context to reference a reason, nor is he merely reporting the fact that he accepts that there is a reason based on authority. If the agent simply did not have propositional attitudes about the elements of a reason, it is non-sensical to claim that he recognized a reason. Just as it would be to claim that Tom recognized his friend Michael in the

- it is not sufficient to recognize the elements, one must also recognize that the reason relation obtains¹⁵⁵;
- it is not necessary to recognize the elements.

Since it is the last point that is critical to the question at hand, it is worthwhile elaborating further. In recognizing a reason, the subsidiary propositional attitudes about the elements play the analogous role to the beliefs about the circumstances and background conditions in the previous example. For any putative case of recognition of a reason this provides a principled means for understanding the demands of the veridical character of recognition, and based on this distinguishing between successful and unsuccessful cases of recognition.

To be clear, an agent can successfully recognize a reason when his beliefs about the elements of the reason are true, and he can successfully recognize a reason when his beliefs about the elements of the reason are false. Further, an agent can fail in recognizing a reason¹⁵⁶ when his beliefs about the element of the reason are true, and he can fail in recognizing a reason when his beliefs about the elements of the reason are false.

crowd, if Tom had no belief about such a person as Michael, or that he was a friend. This would not constitute an act of recognition because the preconditions for such an act do not obtain. To think that an agent can recognize a reason by merely recognizing that the reason relation obtains, but without having propositional attitudes about the elements of the reason is to make too little of recognition.

¹⁵⁵ In just the same way that facts which comprise the elements of a complete reason are distinct from the fact that the reason relation obtains, so too is the recognition of the elements distinct from the recognition of the reason. I.e., if the agent does not recognize that the reason relation obtains, while he may recognize the elements of a reason, he does not recognize the reason itself for he does not recognize that the elements stand in the reason relation.

¹⁵⁶ This locution is intended to pick out cases of unsuccessful recognition, as opposed to merely cases in which an agent fails to recognize a reason through omission, as it were.

To distinguish between successful and unsuccessful cases of recognition, the critical question is not whether the agent's beliefs about the elements of the reason are true, but were his belief true, would they stand in the reason relation to one another, and is this the relation the agent recognizes.

To illustrate the point, it will be helpful to step through an example that may put intuitions to the test. Consider a few slightly regimented variations of the axe murderer example:

Variation 1:¹⁵⁷

- Belief 1: The axe murderer believes that buying the axe is a necessary means to commit the axe murder.
- Belief 2: The axe murderer believes that committing the axe murder would *not* be a valuable outcome.
- Recognition: It is not the case that the axe murderer recognizes that the reason relation obtains between the first and second belief.

In this variation the axe murderer's beliefs are true and he does not recognize that a reason relation obtains between these two beliefs. There is no act of recognition. And from a third person perspective it would be appropriate to say 'The axe murderer did not recognize a reason to buy the axe.'

Variation 2:

- Belief 1: The axe murderer believes that buying the axe is a necessary means to commit the axe murder.

¹⁵⁷ Of course it is possible for an agent to successfully recognize a reason based on true beliefs. This case does not receive explicit treatment here because it is not at issue. However, this should not be read to suggest that such a case is not possible.

- Belief 2: The axe murderer believes that committing the axe murder would *not* be a valuable outcome.
- Recognition: The axe murderer recognizes that the reason relation obtains between the first and second belief.

In this variation the axe murderer's beliefs are true and there is a putative case of recognition. However, the putative case of recognition is unsuccessful because the first and second beliefs do not stand in the reason relation to one another. This is an example in which the demands that stem from the veridical character of recognition are not met, and therefore this is not an example of recognition.¹⁵⁸ And from a third person perspective it may be appropriate to say 'The axe murderer thought he recognized a reason to buy the axe.'

Variation 3:

- Belief 1: The axe murderer believes that buying the axe is a necessary means to commit the axe murder.
- Belief 2: The axe murderer believes that committing the axe murder would be a valuable outcome.¹⁵⁹
- Recognition: The axe murderer recognizes that the reason relation obtains between the first and second belief.

¹⁵⁸ Consider, for a moment, how strange this situation would be. It is not merely the case that the axe murderer has failed in his reasoning, rather it is that reasons themselves do not seem to have requisite traction with the axe murderer. For this reason, it can be suggested that this is not a case in which the agent is irrational, but a case in which the agent is unreasonable. This is deserving of more discussion, but for the moment it is worth pointing out that this is an example of one of two ways in which an agent can be unreasonable. I.e., an agent can be unreasonable by steadfastly seeing a reason relation as obtaining between elements that do not stand in a reason relation, or by being unable to recognize a reason relation as obtaining between elements that do stand in a complete reason relation – given that suitable conditions apply.

¹⁵⁹ The proposition is so repugnant as to make it difficult to entertain the possibility that an agent would genuinely believe it to be true. But, in order to avoid undermining the axe murderer, for present purposes that it is exactly the prospect to entertain, i.e., the axe murderer believes that committing the axe murder would be valuable.

In this variation the axe murderer's second belief is false and there is a putative case of recognition. Despite the fact that it is based on a false belief, the putative case of recognition is successful because were the axe murderer's beliefs true they would stand in the reason relation to one another, and this is the relation that the axe murderer recognizes. This is an example in which the demands that stem from the veridical character of recognition are met, and therefore this is a successful example of recognition. And from a third person perspective it may be appropriate to say 'The axe murderer recognized what he thought was a reason to buy the axe.'

These variations illustrate that the veridical character of recognition is demanding, but not too demanding to play a role in rationality.¹⁶⁰ Paradoxical as it might sound, an agent can recognize a reason to phi, even though he has no such reason. Reflecting on the veridical character of recognition has shown that this is because the relevant question is not whether the agent's beliefs about the elements of a complete reason are true, but if they were true, would they stand in a reason relation.

Before moving on, it is worthwhile to point out that there are some parallels and differences between this view and the well known account of rationality Parfit offered in "Rationality and Reasons"¹⁶¹ and elsewhere.

¹⁶⁰ I.e., despite the fact that the axe murderer had true beliefs in Variation 2, he did not recognize a reason because the reason relation did not obtain between his beliefs; whereas in Variation 3 the axe murderer had a false belief yet nonetheless he was successful in recognizing a reason to buy the axe.

¹⁶¹ 'Rationality and reasons', in Dan Egonsson, Jonas Josefsson, Björn Petterson & Toni Rønnow-Rasmussen (eds.) *Exploring practical philosophy: from action to values*, Aldershot: Ashgate, 2001, pp. 17-39.

Parfit similarly marks a distinction between domains in which the truth or rationality of the underlying beliefs are relevant from those in which they are not.

Leave aside for the moment the fact that Parfit is discussing the rationality of a desire, and the discussion to this point has been concerned with whether or not an agent successfully recognizes a reason – this is a reflection of the different conceptions of practical reasoning being pursued – and the parallels become obvious.

Parfit argued that an agent can form a rational desire based on false or irrational non-normative beliefs. Earlier in this chapter it was argued that an agent could successfully recognize a reason based on false beliefs (the same arguments would hold regarding irrational beliefs) about the elements of the reason.

Further, Parfit argued that an agent could not form a rational desire based on a false or irrational normative belief. In this instance a normative belief could be a belief about whether there is a reason, or for present purposes, whether the reason relation obtains. In the current discussion, it was argued that an agent could not successfully recognize a reason based on a false (or irrational) belief about whether the reason relation obtained.

The parallels are evident, but so are the differences. Perhaps the greatest point of difference between the current account and Parfit's is that on the current account rationality reflects not only the reasons that an agent recognizes, but

also the role this reason plays in the agent's deliberation. Indeed, the current account will suggest that in cases of irrationality there is a tension that emerges between the reasons the agent recognizes and the role those reasons play in deliberation. For Parfit, this second important element is not a focus of discussion. Focusing, as he does, on mental states obscures the question of how the agent wrestles with the reasons he recognizes, which seems the more substantial and interesting aspect of rationality to explore. Unfortunately this seems a general feature of accounts that focus too much on the agent's mental states.

If the discussion to this point shows how the suggested view is similar to, and different from Parfit's view, it also offered an answer to the first objection. In so doing, it also suggests the answer to the second objection, i.e., that the difference between recognition and belief is not significant enough to warrant the propagation of mental states relied on in an account of rationality. For one of the ways beliefs and recognition differ is that beliefs do not share the veridical character of recognition. What then is the relationship between recognition and belief?

While it may be natural to think that recognition implies belief, but belief does not imply recognition, it turns out that recognition and belief are strictly separable. That is to say that an agent can recognize something without believing it, and can believe it without recognizing it. To see this, consider an example of each type of case.

Presumably, to believe that something is a reason, an agent must be familiar with the notion of a reason. Therefore, if a person is not familiar with the notion of a reason, he cannot have a belief that something is a reason. An agent who is unfamiliar with the notion of a reason, a child for instance¹⁶², can nonetheless act for a reason. In order for an agent to act for a reason, he must recognize the reason. That is to say he must recognize that a consideration counts in favor of an action. Therefore, an agent can recognize a reason even if he does not believe that it is a reason.

One can also believe that X is a reason, without recognizing it as a reason. This may be the case, when one takes something on authority or faith. For example, an agent considers x, but does not come to the view¹⁶³ that x is a reason to take an action because he does not understand the relevant considerations. A figure of authority orders the agent to take an action based on x. Even without understand the basis for the reason, the agent comes to believe that x is a reason. This is a case in which the agent believes x is a reason, but does not recognize it as a reason.¹⁶⁴

¹⁶² As Parfit points out the fact that children who are unfamiliar with the notion of a reason can nonetheless act for a reason, counts against the primacy of believing that something is a reason. See: Parfit, *On What Matters*, Oxford University Press, 2011.

¹⁶³ For present purposes, this term is intended to be ambiguous between believing x is a reason and recognizing x as a reason.

¹⁶⁴ While this is a case in which an agent believes there is a reason without recognizing a reason, the presence of the second-order reason from authority may raise doubts. A brief discussion of this complexity should allay concerns.

According to the position advanced so far it might seem that if the agent takes the action in question for a reason, then he recognizes a reason. Why does this not argue against the conclusion just offered? Succinctly put because in this case the agent recognizes the second-order reason and forms a beliefs about the first order-reason, he does not recognize the first-order reason. As a result this is still an example of the fact that an agent can believe that x is a reason, without recognizing that x is a reason.

Further support for this conclusion can be generated by reflecting on which considerations would change the agent's assessment of the reasons he has. For example some reasons have cancelling conditions, i.e., conditions which cancel the reason. Reasons based on promises are of this kind. When the promised party releases the promising party from carrying out the promise, the promising party no longer has reasons based on the promise for carrying it out.

As these examples illustrate an agent can recognize a reason without believing he has one, and can believe he has a reason without recognizing a reason. And this answers the second objection. While belief and recognition may be connected in many instances, they remain quite distinct.

The distinctive character of recognition importantly helps answer a criticism leveled at externalist accounts like the one presented here. Christine Korsgaard puts forward the following poignant version of this criticism in “Constitution of Agency”¹⁶⁵: according to realism “there are facts, which exist independently of the person’s mind, about what there is reason to do; rationality consists in conforming one’s conduct to those reasons... we must explain why the person finds it *necessary* to act on those normative facts, or what is it about

Based on this, consider an elaboration of the current example. An army Colonel promises an influential politician that his soldier, the agent in question, will clean the politician’s property. Not understanding promises very well, the soldier does not recognize that this promise is a reason to clean the politician’s property. Ordered by the Colonel to clean the property because of the promise, the soldier comes to believe that the promise is a reason. In time, the soldier comes to learn that fearing scandal the politician released the Colonel from his promise. This does not change the soldier’s assessments of the reasons he has, which are after all based on the orders he received from the Colonel. And indeed the soldier may be well justified. For the Colonel may insist that whether or not he was released from his promise, he had made a promise and intends that it be carried out. To the extent that the Colonel has authority, the soldier still has reason to clean the politician’s house. Learning of cancelling conditions of the first order reason may not change the agent’s assessment of the reasons he has, for while he believes he has a first order reason, the reason he recognizes is a second order reason.

Authority is also subject to limits and canceling conditions. Alternatively, if the agent learns of cancelling conditions for the second-order reason, he may change his assessment of the reasons he has. For example, if the soldier comes to believe that the Colonel’s order constitutes an abuse of his power and is beyond the limits of his authority or cancels his authority altogether, the soldier may well reconsider his reason for cleaning the politician’s property; this despite the independent merits of cleaning the politician’s property or fulfilling a promise. Learning of cancelling conditions of the second-order reasons may change the agent’s assessment of the reasons he has. For while he believes he has a first order reason, the reason he recognizes is not a first order reason, but a second order reason. I.e., the soldier believes he has a reason, but does not recognize it as a reason.

¹⁶⁵ Korsgaard, Christine, Constitution of Agency, Oxford University Press, 2008.

her that makes them normative *for her*. We must explain how these reasons get a grip on the agent.”¹⁶⁶

For Korsgaard, in order for the reasons to ‘get a grip on the agent’ two things must take place. First, the reasons must be capable of motivating the agent. And second, they must be capable of motivating the agent as a guide or a norm, which entails it must be possible for the agent not to be motivated by the reason.

Internalists accounts, Korsgaard argues, meet the first of these requirements, but they do so by definition. As a result, it is not possible for the agent to have a reason that he is not motivated by. And therefore, internalist accounts are not able to meet the second of these requirements.

On the other hand externalist accounts, Korsgaard argues, are able to meet the second criteria, for while external reasons offer norms, it is possible for agents to fail to be motivated by external reasons. However, externalist accounts have trouble explaining force of reasons on the agent.

In a move that is amicable to the account offered here, Korsgaard offers the following challenge ‘if what we mean when we say that the person is caused to act by his *recognition* of certain considerations as reasons, then we must say *what it is* that he recognizes.’¹⁶⁷

¹⁶⁶ Korsgaard, Christine, Constitution of Agency, Oxford University Press, 2008, pg 52,3.

¹⁶⁷ Korsgaard, Christine, Constitution of Agency, Oxford University Press, 2008, pg 56.

This is, in part, the question that will be taken up in the next section as the discussion explores the unique character of recognition, how recognition differs from belief, and what makes it well suited to play this central role in an account of rationality.

Consider two statements:

- 'I recognize the truth of that statement'
- 'I recognize the error of my ways'

Each statement seems to have more force than simply saying:

- 'I believe that statement is true'
- 'I believe that my ways were in error'
 - Or in more common parlance 'I believe I was wrong'

Plausibly the different force of the first statements stems from the force of recognition in deliberation. To speak loosely, that which one recognizes does not stand as one proposition among many. It stands with a different level of surety. Possibly this stems from the veridical character of recognition. From the third-person perspective it has already been noted that this veridical character has limits; and recognition should not be taken to be a mark of truth. But things differ from the first person perspective. The agent is not only inclined to hold as true that which he recognizes, but it is difficult to not do so. But the force of recognition seems to be more than this.

There is a sense in which recognition harkens back to something already established.¹⁶⁸ In recognizing something, there is a sense in which the agent re-cognizes it; brings it before their mental faculty, as it were, anew in relation to something already established. In so doing the force of that which was already established is conveyed to that which is recognized.

Agents recognize many things, family members, old neighbourhoods, and the error of one's ways. In recognizing certain kinds of things, it is easier to see the force of recognition, as it is with the case of recognizing an old neighbourhood or the error of one's ways. For some, part of what it is to recognize one's old neighbourhood is to re-cognize it as a place with the emotional attachments associated with it. Part of what it is to recognize the error of one's ways is to re-cognize one's actions as flawed in important aspects. This re-cognizing forcefully changes one's conception of that which is recognized.

This characteristic of recognition has an important role in rationality. For an agent, learning that he has the ability to make choices, that his actions can affect the world and bring about different outcomes, that there are facts that have bearing on the choices he (should) make, is part of his development. It is part of his development as an agent. This is part of what it means to become a creature with reason. This is part of what it means to become a rational creature. For an agent to recognize that considerations bear on his choices and the action he (should) take is part of what it is to recognize something as a

¹⁶⁸ For example an agent does not recognize someone he meets for the first time, whereas he can recognize someone he already knows. Or if he does recognize someone totally new it may be in the sense in which there is a reference to something already established, such as when on meeting someone new the agent recognizes the person as someone with whom he shares a common friend or as a person of a certain political persuasion. This need not only happen with something new. It can also happen with something long familiar such as when an agent recognizes their sentimental attachment to an old desk.

reason. Roughly put in short form, when an agent recognizes that something is a reason for him, he re-cognizes it as pertinent to him and to his making of a choice as a creature of reason. When he recognizes something as a reason, an agent recognizes that it has a role to play in deliberation. This is perhaps the most poignant sense in which recognition differs from belief; it is the means by which reasons 'get a grip on the agent'. It is in virtue of this that recognition has a special role to play in rationality. And it is in this sense that reason is directive when an agent recognizes a reason as a reason.

This is an important part of the story, but it is only a partial description. For if this is all reason involved, then reason would be impotent. Reason would be directive without being efficacious. It is not enough for an agent to recognize that a reason has a role to play in deliberation, the reason also needs to play that role, the appropriate role, in the agent's deliberation. That is to say, the agent needs to take the reason to be a reason. Or, more specifically the agent needs to take the reason to be the reason he recognized it to be. This is the topic to which the discussion will turn its attention next.

Taking something as a reason

To explore the possibility that rationality involves taking a reason to be a reason, it will be helpful to address the several faces of a poignant challenge before moving on to offer reflections of the nature of deliberation.

While the challenge may take many forms, the basic idea can be expressed simply: there is reason to think deliberation does not play the central role the suggested account describes.

First, consider the case of the agent reaching to catch the glass as it falls. In many such cases the agent does not actively deliberate, he simply reaches for the glass. While this may have been in accord with the agent's reasons and even with his recognized reasons, it is not a case in which a recognized reason played a role in deliberation.

Cases of this kind do not present a counter-example to the suggested view, but point to the limits of the account. Cases of this kind are cases in which an agent acts, but his act is not an action. It is merely a reflex. The current account is concerned with intentional action, that is, action for a reason. Reflexes are beyond the scope of the current account.

Even so, examples which show that action is removed from deliberation can cast doubt on the centrality of deliberation.

For example, consider the agent who glances at his phone while waiting in line. This can occur as the result of active deliberation, but can also occur without it. In some instances it is not an action the agent decides to undertake, it is just something the agent does. It is a habitual action, i.e., it is a result of habit and it is an action. That is to say, it is intentional under some description. The agent does them for a reason, even if not directly as the result of active deliberation.

This may seem to raise challenges for the suggested account and the emphasis it places on deliberation. There is much to explore in cases of this kind and the role of habits in general; too much to be treated sufficiently in the current effort.

Nonetheless there are a few things about the nature of deliberation and characteristics of rationality that examples of these kinds bring to light.

First of all it raises the need to distinguish between active deliberation and deliberation broadly defined. Some deliberation is active in the sense that it is front of mind; i.e., the agent is aware of the deliberation.¹⁶⁹ A quintessential case of such active or foreground deliberation is when an agent sets out to arrive at a decision. Not all deliberation takes this form. Perhaps most of it does not. In many instances, in making a decision the agent is not even aware of his deliberation. Indeed the situation is even more poignant. Much of an agent's deliberation is not immediately accessible or transparent. Much of an agent's deliberation happens in the background. The distinctive character of introspection notwithstanding, there are many cases in which an agent learns the outcome of his own deliberation through his choices or retrospectively. This is reflected in the colorful advice sometimes offered to agents consciously wrestling with large decisions. 'Flip a coin' the advice suggests 'and as you see the result you will know what you really think'.

This kind of deliberation is quite different than the conscious calculating of pros and cons that is so often the caricature of rational thinking. This is the kind of reasoning in which agents respond to the often substantive reasons that they have through a host of their rational faculties, including their emotions. Rather than a distinctive activity engaged in on rare occasions, deliberation broadly defined is, for many, an ongoing condition of conscious existence.

¹⁶⁹ This is similar to the distinction between foreground and background introduced by Pettit and Smith in "Backgrounding Desire". Pettit, P., Smith, M. "Backgrounding Desire", *The Philosophical Review*, Vol. 99, No. 4 (Oct., 1990), pp. 565-592.

This initial clarification of the relevant definition of deliberation may go some way to alleviating concerns about the centrality of deliberation in the current account. For the claim at hand is quite different than the straw man view that agents actively deliberate on the basis of each reason they recognize.

Nonetheless, even this weaker view may seem too strong in considering examples of habitual action, and this is for good reason. Agents have limited rational capacities; and one role of habits is to economize the deliberation required. Habits play this role by bundling actions; and agents often perform bundled actions together in relevantly similar circumstances. For example, consider the case of the agent who habitually puts on his seat belt and checks his mirror every time he sits in the driver seat with the intention of driving the car. The habit of doing this saves the agent the expense of deliberating about whether or how to take these steps each time he gets in the car. In this way habits economize the deliberation required.

Importantly however, habits themselves are responsive to reason and deliberation – some more immediately than others.¹⁷⁰ For example the breaking old habits and forming new ones is possible through conscious effort, though difficult at times. Interestingly, habits also demonstrate that the relationship between deliberation and the action which results from it can be quite temporally extended. So much so that habits can outlive the reasons for which they were formed. For example, it is a common experience as one proceeds through life that the habits one formed for good reason at one stage in life no longer serve him well as he enters different stages in his life. Habitual

¹⁷⁰ Certainly the link between habits and reason is not entirely straightforward and is deserving of more attention than it will receive here.

actions, like those of the driver, are not examples of actions without deliberation. However they do illustrate the fact that at times the relationship between deliberation and action is quite extended.

There are other examples in which it can be quite difficult to see the connection between an agent's action and his deliberations. For example, when getting up from his seat an agent may take his first step with his right foot or his left.

Taking a step is an intentional action, though in the majority of ordinary cases it is difficult to see that the agent had a reason to take a step with one foot rather than another, let alone that he would deliberate about which foot to step with.

This, however, is not an example of an agent acting intentionally in the absence of deliberation. It is an illustration of the saliency of the level of descriptions.

While the agent may not have reason to take his first step with his left foot rather than his right foot, he may well have reason to cross the room, to go to the store or to get food. Examples of this sort are something of a red herring.

They do not provide distinct basis to be concerned that the current account calls for deliberation to play too central a role, for they are merely a description of common cases of practical reasoning at a level of description which obscures the relationship between reasons, deliberation and action.

Taken together, these discussions of cases in which deliberation is not closely connected to action have hopefully assuaged the concern about the central role of deliberation on this account. Doubtless it would be helpful to have a clear account of deliberation to make this point more clearly. Unfortunately offering such an account of deliberation is beyond the current scope. Deliberation is

complex. It is properly the subject of a field or several fields of study, not the subject of a section of a chapter. Nonetheless a sense of the complexity of deliberation can be offered by reflecting on the different questions to consider. This following discussion will seek to touch on aspects of deliberation at a basic level while drawing out several salient aspects that touch on the current account.

There are a number of ways that one can carve up the vast terrain. One simple, but fruitful way of doing so is to note that certain characteristics of deliberation have to do with what might be called the logic of reasons, whereas other characteristics have to do with the fact that it is agents that deliberate. Implications of each of these will be looked at in turn.

To avoid explaining the obscure with reference to the more obscure, a few words should be said about the logic of reasons. This is not a term of art, but a descriptive phrase. Reasons, by and large, are prima facie reasons. That is to say, though an agent may have a reason to take an action, it may not be the case that he ought to take that action all things considered, because he may for instance have more compelling reasons not to take the action or to take another action. Here the 'logic of reasons' will be taken to mean the logic according to which the fact that an agent has a reason becomes the fact that an agent ought to do something all things considered.

As a subject, the logic of reasons has yet to receive sufficient treatment. Some areas, such as questions pertaining to instrumental reasons, have received more attention than others. For example, if one action is far more likely to bring

about the same positive outcome for no greater cost than another action, then all other things being equal as far as these instrumental reasons are concerned, the agent has more reason to perform the first action than the second. But such idealized comparisons of instrumental reasons are only a small fraction of the relationships between reasons.

At times reasons compete. The example of instrumental reasons just mentioned is an example of this. It may also be the case that an agent has reason to bring about one valuable outcome, and reason to bring about another valuable outcome. And to complicate matters, it may be the case that one of the outcomes is more valuable than the other, or it might be the case that the value of the outcomes is not commensurable. Indeed, it may not even be comparable; or may only be incompletely comparable. The question of how to tally up the competition between such reasons is deserving of attention in its own right.

Further, there may be question of whom the outcome is valuable for. I.e., the action an agent can take may be valuable for one person, while another action the agent can take would be valuable for another person. Depending on the circumstances, it can be a substantive moral question of how these reasons compete.

And the complexities can multiply. For example, one and the same action may be good for some and bad for others. Similarly it can be a substantive moral question to determine how the various effects should be tallied.

Not all questions verge so closely to the terrain of value theory. Some questions stem from the variety of relationship between reasons themselves. For example, some have suggested a specification relation exists between reasons. I.e., the reason one has for going to the train station may be a specification of the reasons one has for keeping one's promises of picking up a friend from the train station.

As was discussed earlier, there are also second order reasons. And some reasons may admit to a number of different cancelling conditions. And admittedly this is likely only a sampling of the reasons that there are.

The complications are multiplied by the fact that deliberation is done by agents. As such, deliberation is not merely a matter of working through the logic of reasons. Deliberation is in part a creative activity¹⁷¹, and can reflect the values and character of the agent deliberating. Further, agents have a variety of abilities, and are severely limited in the best of cases. Assessing each of the reasons that pertain to agents during the course of a normal day is beyond the realm of possibilities for agents.

As one would expect, deliberation is complex, intricate and sophisticated. While it would be helpful to have a clear prescription for what it would mean for a reason to play an appropriate role in deliberation, it is unlikely that any easy formulaic response will suffice. Indeed there are cases in which it is appropriate for a recognized reason to play no role in deliberation.

¹⁷¹ For example, what one considers while deliberating can be reflective of the agent's intelligence and imagination.

Consider a few cases in which a reason that an agent recognized does not play a role in this deliberation. For example, consider the case of the agent who knew that in order to fulfill his promise and pick up his friend at the train station, he would need to go to the train station. The agent recognized that he had a reason to go to the train station to pick up his friend. This reason can fail to play a role in deliberation for several different reasons.

This might happen as the result of a failure of some sort. For example, the agent may have forgotten his promise to his friend, or the fact that in order fulfill his promise he needed to go to the train station, or that his friend was arriving this afternoon. However, the omission of the recognized reason from deliberation in this case is not a failure in deliberation, but a failure in the precondition for it. Indeed, one might wonder whether it is apt to describe an agent in this position as recognizing a reason. It may be the case that he had recognized the reason, but no longer does. Therefore, not including this reason in deliberation may have been a failure, it is not a failure to take as a reason that which he recognized to be one.

In addition, there may be many circumstances in which the question of going to the train station never came up.¹⁷² For example, the agent's friend may not be arriving until some time in the distant future. Given this, there is no call for an agent to deliberate about whether or not to go to the station at this point in time. As such, if there is no relevant deliberation¹⁷³, then the agent has not failed to

¹⁷² Care needs to be taken here, for in some cases not deliberating about a possible action can itself constitute not taking as a reason that which one recognized to be a reason.

¹⁷³ And the fact that there is no deliberation does not in itself constitute a failure on the part of the agent to take as a reason that which he recognized to be one.

take as a reason that which he recognized to be one, even if he did not deliberate on the basis of the reason he recognized.

This points to an important aspect of what it is to give a reason an appropriate role in deliberation which is made more explicit by another example in which an agent does not fail to take a reason as a reason even if he does not deliberate on the basis of the reason he recognized. Consider the example of the parent taking care of his child. The parent knows that the law requires him to take care of his child, and recognizes that this is as a reason to take care of his child. However, in deliberating about how to act, the parent does not deliberate on the basis of this reason. He simply takes care of his child because he is his child; or perhaps more aptly put because the child needs to be taken care of. The fact that the law requires it never crosses his mind. In this case, the parent's actions are overdetermined – normatively overdetermined. He has an overabundance of reasons to take care of his child. He does not need to deliberate on the basis of each of these reasons in order to give each of these reasons its appropriate role in deliberation. For in recognizing a reason, an agent recognizes that a consideration counts in favor of a proposition or an action. In cases in which the agent's action is normatively overdetermined, some of the reasons in favor of the action simply do not add more support for taking the action. They have no possibility of affecting the agent's balance of reasons, and therefore need not play a role in deliberation. To give a reason its appropriate role in deliberation in part is to give it a role in deliberation if it has the possibility of affecting the balance of reasons.¹⁷⁴ In other words, to take as a reason that which one recognized to be a reason is to incorporate that reason

¹⁷⁴ This also means that if there is overwhelming contrary reasons, it may be the case that an agent can fail to take a reason into consideration in deliberation without thereby failing to give the reason an appropriate role in deliberation.

in deliberation if doing so has the possibility of effecting the agent's balance of reasons.

This last example is revealing. It points to the fact that playing an appropriate role in deliberation has to do with affecting the agent's balance of reasons. This is not to say that this is what constitutes playing an appropriate role in deliberation. As was just illustrated, what role is appropriate for a reason to play will depend on the reason itself as well as the other reasons that pertain to the question at hand. Nonetheless, this observation does present a standard that can be used to determine whether a given reason played an appropriate role in deliberation. If a recognized reason would affect the balance of reason if it were to play an active role in deliberation then in order for the reason to play an appropriate role in deliberation, it needs to play an active role in deliberation.¹⁷⁵ Needless to say, given that this pertains to deliberation broadly defined, this is not a criteria which can easily be applied since the agent himself may not be aware of the role a given reason played. The real benefit of the criteria is a degree of conceptual clarity.

Taken together an intricate and nuanced picture of rationality begins to emerge. This befits the complexities agents grapple with as they navigate a world with a multitude of possibilities that requires the ability to perceive alternatives and ultimately make choices. As agents learn about themselves and the world around them they develop the ability to recognize that considerations bear on their choices. They develop the ability to recognize reasons. But this is the

¹⁷⁵ This also suggests that there is a degree of asymmetry to which reasons need to play an active role in deliberation in order to play an appropriate role in deliberation. I.e., recognized reasons that count against the recommendation of the balance of reasons need to have played an active role in deliberation in order for the agent to have deliberated appropriately. This same is not the case for reasons that count in favor of the recommendation of the balance of reasons.

beginning and not the end of the exercise of their rational faculties. Agents must also take the reason to be a reason. For in order to choose agents must deliberate amongst a multitude of such considerations which can have a complex relationship with one another, and agents must do all this with very limited resources in an on-going basis. There are a variety of standards for success in how agent deliberate amongst the reasons they have. Part of this depends on the logic of reasons, and part on the character of choice and deliberation as creative acts. For example, the results of an agent's deliberation can be in line with the balance of his reason, or in line with the balance of his recognized reasons; his choice could be the virtuous choices, loyal choice or creative choice. A full description of the variety of possibilities would likely be encyclopedic. Nonetheless in reflecting on rationality we can gain an understanding of these two distinct but important components of rationality. I.e., reason is directive and efficacious. Reason is directive when an agent recognizes something as a reason; and reason is efficacious when an agent takes a reason to be a reason.

Stage 3: Irrationality

With this understanding of the significance of directive and efficacious characteristics of rationality, the discussion can again turn to irrationality.

There are a number of necessary conditions for an agent to be rational. While it may be true that a failure to meet any of these conditions constitutes a failing of rationality, it is not true that each of these failings constitutes irrationality.

Irrationality is a failing of a specific kind.

For instance, arguably consciousness is a necessary condition for rationality, however an agent lacking in consciousness is not irrational, but non-rational. Consciousness, one may think, is a precondition of rationality. And, it seems reasonable to say that if an agent fails to be rational because she lacks a precondition for rationality, she is not irrational, but non-rational.

A unique aspect of the two conditions that have been discussed so far is that a failure with regard to these conditions is a case of irrationality.

Given that there are two conditions, they can be present in 4 combinations.

1. P *recognizes* X as a reason, and *takes* X to be a reason.
2. P does not *recognize* X as a reason and does not *take* X to be a reason.
3. P does not *recognize* X as a reason, yet *takes* X to be a reason.
4. P *recognizes* X as a reason, but does not *take* X to be a reason.

Case (1) describes a situation in which two necessary conditions for rationality are met.

Case (2) describes a situation in which two necessary conditions for rationality are apparently not met. But this does not argue against their being necessary conditions, but reflects the relation between them. A more explicit statement of the necessary conditions is: P is rational only if (P recognizes X as a reason, if and only if he takes X to be the reason he recognizes it to be). Understood in this way, Case (2) describes a situation in which two necessary conditions for rationality are met.

Case (3) describes a failure on the part of the agent, but this is a failure of a very different kind. Consider for a moment what such a case would be like. The agent takes as a reason that which he does not recognize to be one. Though the agent does not see the consideration as directive, it is efficacious in his deliberation. This is not a case where the agent fails to be ruled by reason, but it is a case where he fails to be ruled by his reason. This is not a case of irrationality, but a case of non-rationality. This will not be argued further here. The best argument for it seems to be considering how strange it would be for a person to be in this situation. But if some think it is better to describe this as a special kind of irrationality, nothing in this discussion should turn on that.

Case (4) describes a particular failure on the part of the agent. The agent's reason is directive, but not efficacious. Consider an example that would be of this kind. I.e., an example in which the agent did not meet the criteria described above. For example, imagine if the agent in question did not go to the station because when he deliberated about whether or not to go to the train station, he never considered the reason he had to go in order to pick up his friend from the station. It is not that something intervened and he no longer recognized the reason that he had, nor that he had more reason not to go. He simply failed to take the fact that he made a promise to be a reason; he simply failed to give it the appropriate role in deliberation. If the agent recognized the reason, did not forget it, deliberated about the possibility, but did not give the reason its appropriate role in deliberation, his choice is in a way inexplicable. There is no reason for it, in part because it is a failure of reason. Cases of this kind, it is

suggested are cases of irrationality. Rather than merely a failure of reason to rule, cases of irrationality constitute a failure *in* reason's ruling.¹⁷⁶

Descriptive Requirement

As a suggested account of irrationality, it is appropriate to ask whether this account meets the Descriptive Requirement articulated earlier. I.e., will it offer a descriptive account of the failure of irrationality that is common to cases of irrationality. The suggestion at hand is that: cases of irrationality are cases in which an agent *P* *recognizes* *X* to be a reason, but does not *take* *X* to be the reason he recognizes it to be.

To completely satisfy the first requirement, one must also show that this failing is common to all cases of irrationality. This is too ambitious for present purposes, but some progress can be made towards this goal by showing that the suggested view can accommodate Davidson's descriptive account and serve as the basis for Broome-style rational requirements. The hope is that by showing that the suggested view is compatible with the successful elements of each of these views, it will countenance its plausibility.

One obvious point to address is that Broome-style rational requirements look very different than the current suggested view. In this section, it will be argued that the current view is similar to and different from Broome-style rational requirements in important respects.

¹⁷⁶ To put this more explicitly, reason can fail to rule (i.e., an agent can fail to be rational) in two ways, reason can not rule (i.e., agents can be non-rational) or reason can fail in ruling (i.e., an agent can be irrational).

To see this it will be helpful to briefly look at a debate between Broome and Kolodny about rational requirements that has received considerable attention. In “Rationality and Reasons”¹⁷⁷, Kolodny argues that, contrary to Broome, at least some rational requirements are narrow scope.¹⁷⁸ Kolodny’s main argument is essentially comprised of two different claims: first, that at least some rational requirements are process-requirements; and second, that at least some of these process-requirements are narrow-scope.

Kolodny distinguishes between “‘state-requirements,’ which simply ban states in which one has conflicting attitudes, and ‘process-requirements,’ which say how, going forward, one is to form, retain or revise one’s attitudes so as to avoid or escape such conflict-states.”¹⁷⁹ And he stipulates that “Any account of rationality that aims to capture our ordinary attributions cannot consist solely of

¹⁷⁷ Kolodny, “Rationality and Reasons”.

¹⁷⁸ Note that the reason this seemingly esoteric topic has received so much attention is because of its significance to the normative question. I.e., if it is the case that the requires relation (and rationality) is normative and rational requirements are narrow scope, then an agent can seemingly bootstrap himself into being normatively required to have a belief, preference, intention, etc. Or conversely, since many seek to avoid this bootstrapping conclusion, if one can show that some rational requirements are narrow scope, one can pose a dilemma for one’s opponents by showing that they must either accept the bootstrapping conclusion or give up the claim that requires relation (and rationality) is normative.

While this is a reason many have closely followed this debate, it is not a topic of particular concern here for two reasons. The first has already been discussed in several different places in the current work. Specifically, it has been argued that it is not sensible to ask whether rationality is normative simpliciter. Rather, it is more fruitful to explore whether a given conception of rationality (normative or not) is feasible given a broader conception of practical reasoning. This is a reflection of the extent to which concepts within a given conception of practical reasoning are inter-defined, as discussed extensively in Chapter 1: Introduction.

The second reason this debate is not a particular concern for present purposes is that, as will be argued shortly, this debate focuses on a questionable characteristic of rational requirements.

For reasons that will become apparent later in this dissertation, it is also worthwhile to note that while it is problematic to claim that rationality is normative and rational requirements are narrow-scope, there is no corresponding difficulty in claiming that rationality is not normative and rational requirements are wide-scope.

¹⁷⁹ Kolodny, “Rationality and Reasons”, pg 6.

state-requirements, which say “Not that conflict!” It must consist, at least in part, of process-requirements, which say “Do this to avoid or resolve that conflict!”¹⁸⁰

Based on this stipulation, he then develops the “Reasoning Test” to explore whether a process-requirement is wide or narrow scope: “The process-requirement governing conflict between A and B is wide scope – i.e., one is rationality required (either not to have A, or not to have B) – only if, from a state in which one has conflicting attitudes A and B, (i) one can reason from the content of A to dropping B and (ii) one can reason from the content of B to dropping A.”¹⁸¹

And then he proceeds to show that a number of putative wide-scope requirements fail to meet the Reasoning Test. For example, “I-WS: Rationality requires one (either not to believe that one lacks sufficient reason to X, or not to intend to X).”¹⁸² As Kolodny points out, one can reason from the contents of one’s belief that one lacks sufficient reason to X to dropping of one’s intention to X; but one cannot reason from the content of one’s intention to X, to the belief that one has sufficient reason to X. One cannot reason upstream, as it were from ‘one’s attitudes to a reassessment of one’s reasons for those same attitudes’.¹⁸³ Doing so would not be a case of reasoning but something else, possibly something like wishful thinking or self-deception. As a result, this example fails the Reasoning Test, and should instead be considered a narrow-

¹⁸⁰ Kolodny, “Rationality and Reasons”, pg 6,7.

¹⁸¹ Kolodny, “Rationality and Reasons”, pg 9.

¹⁸² Kolodny, “Rationality and Reasons”, pg 10.

¹⁸³ Paraphrase of Kolodny, “Rationality and Reasons”, pg 11.

scope requirement or as Kolodny puts it: "I-NS: If one believes that one lacks sufficient reason to X, then rationality requires one not to intend to X."¹⁸⁴

Since the suggested view endorses some of Broome's characterization of rational requirements as wide-scope requirements, it must be at odds with some of Kolodny's account, but to be clear there are also points in which the suggested view agrees with Kolodny's view. For example, the suggested view accepts Kolodny's suggestion that rationality is at least in part concerned with how one progresses - by reasoning - from one state to another. This seems intuitively compelling, and the suggested view has no issue with the possible centrality of process-requirements as such, even if it does differ in the conception of those requirements. Further Kolodny's characterization of upstream reasoning as a kind of wishful thinking seems to aptly capture the strangeness of progressing in that fashion. Given these points of agreement, the question is how does the suggested view resist the force of Kolodny's argument?

The answer is relatively straightforward. First, it will be argued that Kolodny's Reasoning Test reflects an artificial simplification of reasoning which is unwarranted. And, second it will be suggested that it will be more fruitful to think of rational requirements in a somewhat different way than either Kolodny or Broome do. These points will be taken up sequentially.

To see this, it will be helpful to revisit the rational requirements Kolodny offered to serve as the basis for his argument that some rational requirements are

¹⁸⁴ Kolodny, "Rationality and Reasons", pg 11.

narrow scope process requirements. Specifically, “I-: Rationality requires one not to intend to X, if one believes that one lacks sufficient reason to X.”¹⁸⁵

The first difficulty is that I- is not obviously a rational requirement. It is not irrational to form an intention merely based on whim or whimsy despite believing that one lacks sufficient reason. Kolodny discusses just such a concern in a footnote that attends the introduction of I-. And he suggests that in such cases, I- is trivially satisfied for in such cases one believes that one does not need any reason to X. But this clearly goes too far. For one can act on whim or whimsy without needing to form a belief about what reasons one would need in order to X. Indeed, such consideration of the matter would seem to fly in the face of exactly what it is to act on a whim or whimsy.

While this characterization of I- may be reflective of the overall character of Kolodny’s treatment of reasoning that will be objected to, it is not itself the point at issue. So for present purposes, a variation of I- can be offered that may be less objectionable. Consider Ia: In forming an intention to F and G, rationality requires one not intend to G if one judges oneself to have more reason to F.

Consider an agent in violation of this requirement. For example an agent who judges himself to have more reason to F rather than G, but forms an intention to G rather than an intention to F. Such an agent would be akratic, and such an agent’s intention would be irrational.

If Ia is a wide-scope process-requirement, according to Kolodny’s Reasoning Test, then it must be the case that the agent can (i) drop the intention to G

¹⁸⁵ Kolodny, “Rationality and Reasons”, pg 10.

based on his judgment that he has more reason to F, and (ii) revise his judgment that he has more reason to F based on his intention to G. But the latter would clearly be a case of impossible upstream reasoning that was discussed above. And so, it appears Kolodny's argument applies to Ia as neatly as it did to I-.

The question is why is the Reasoning Test so restrictive. An agent aware of having an intention that is at odds with his better judgment may well revisit his better judgment, but, as Kolodny points out, it would be wishful thinking if he did so based on his intention. On the other hand, it would be perfectly rational for the agent to revise his judgment about the balance of reasons when he revisits the reasons for his intention. An agent reconsidering the reasons on which he formed his intention may well arrive at a different conclusion about the balance of reasons. For example, in this case, when the agent revisits the reasons on which he formed the intention to G he may conclude that he has more reasons to G than F. In this way, when an agent finds himself in a situation which violates Ia, the agent can reason his way out of this conflict by either changing his intention or by changing his judgment about the balance of reasons. Contrary to Kolodny's suggestion, Ia is a wide-scope process-requirement.

This conclusion would seem to indicate that the suggested view sides with Broome rather than Kolodny, but there is more going on here that is worth noting. Specifically, it will be helpful to make four points.

First, similar to Kolodny, on the suggested view rationality is concerned with how one progresses from one state to another, that is to say, how one reasons.

In this sense, the two views share an orientation. However, they differ in the conception of that process. The Reasoning Test reflects a narrow and prescribed conception of practical reasoning, whereas the suggested view starts with the premise that reasoning rationally can be an intricate, iterative and nuanced activity.

Second, the suggested view differs from both Broome and Kolodny's views in its focus on the reasons an agent recognizes, rather than the agent's mental states, such as beliefs and intentions. This is the basis for the argument that Kolodny's Reasoning Test is too restrictive. I.e., Kolodny is correct to point out that an agent's intention to G may provide that agent with no reason to conclude that he ought to G, or that he has sufficient reason to G. But if the agent rationally formed his intention to G, then he did so on the basis of recognized reasons. And those reasons may well provide him with the basis to conclude that he ought to G, or at least has sufficient reason to do so.

This difference in focus is more important than it may at first appear. As was already discussed, the propositional content of a mental state, and the mental state itself for that matter, may stand in an infinite number of different reasons relations. And an agent may recognize that it stands in some of those relations, none of those relations, or he may mistakenly think that it stands in some other relations. To focus on the agent's beliefs and intentions, without also focusing on the reasons he recognizes is to obscure from vision the basis on which the agent reasons. And as a result, attempts to enumerate rational requirements with this focus will often be stilted, over-prescriptive, and have a seemingly ad-

hoc character dealing with specific characteristics of different circumstances in elaborate detail.

This would seem to suggest that on the current view one might think that there is little virtue in Broome or Kolodny style rational requirements. This is not the case. It is just that one understands the significance of each view quite differently. This brings us to the third point. The debate about whether rational requirements are wide-scope or narrow-scope generally pertains to what might be considered the imperative to remedy an irrationality. That is to say, given a circumstance in which an agent is irrational, say by having an akratic intention, the question seems to be whether the agent can cure or remedy the irrationality in more than one way. The setting for this argument seems to be the implicit shared view that rationality requires one to cure or remedy cases of irrationality.

While the current view shares many aspects with both Broome and Kolodny's views, it does not share either views emphasis on the question of how one can cure or remedy cases of irrationality. In some cases rationality may require an agent to remedy an irrationality, but in other cases – just like violations of other types of requirements - it may not. Rather the focus on rational requirements in the current account is quite different and it sheds a different light on the significance of both Kolodny and Broome-style rational requirements. To see this, it would be helpful to discuss a specific example, taking the opportunity to point out the places where the suggested view agrees and disagrees with Kolodny and Broome.

As was already mentioned the suggested view holds that rationality concerns reasoning correctly, in a specific sense. This seems relatively close to Kolodny's emphasis on process-requirements, though with two specific differences: first, the suggested view is concerned with the reasons an agent recognizes rather than the general content of his mental states, and second, that the suggested view concerns an understanding of reasoning which it would be difficult to narrowly prescribe in the way Kolodny does. The suggested view of rationality can be codified in a fundamental requirement of rationality that is reminiscent of Kolodny-style process requirements, i.e., Rationality requires that if an agent recognizes something to be a reason, then he takes it to be the reason that he recognizes it to be. An agent's failure to meet this requirement of rationality makes it the case that the agent is irrational.

The claim is that this fundamental requirement of rationality is violated in all the canonical cases of irrationality. And, since Broome-style rational requirements can be articulated to codify these canonical cases of irrationality, these Broome-style rational requirements codify situations in which this fundamental requirement would be violated. To see how this might work, let's suggest a fifth rational requirement that is similar to Broome's fourth requirement.

- Fifth requirement: Rationality requires of you that, if you are choosing between F and G and you believe you ought to F, then you intend to F, rather than G.

Given the difference that has already been discussed between recognizing a reason and believing there is a reason, it will be helpful to offer a minor variation to clarify the role played by the outcomes of the agent's deliberation.

- Fifth requirement': Rationality requires of you that, if you are choosing between F and G and all things considered you judge that you ought to F, then you intend F, rather than G.¹⁸⁶

Consider a violation of this requirement, a case in which the agent intends to G. As Broome-style rational requirements are supposed to be, this requirement is strict. Any agent who violates the requirement in this way would be irrational. It is harder to see why one should think this requirement is broad. It appears that the only way an agent can remedy this violation of the requirement would be to form the intention to F based on all the relevant reasons. But this is too quick. The agent can also change his judgment about what he has most reason to do all things considered. A look at the connection between the suggested view and this requirement can shed light on why this is the case.

Recall that in cases of akrasia an agent forms an ATC-J based on r1 and r2, but forms his SP-J based on r1, for example. In forming his intention on this subset of reasons, the agent fails to take r2 to be the reason that he recognizes it to be and is therefore irrational. Such an agent can remedy this failing by forming an intention based on r1 and r2. Alternatively he may retain his intention but cease to recognize r2¹⁸⁷ as a reason. In so doing, the agent would have remedied his violation; and so the suggested view can accommodate both the strict and broad character of Broome-style rational requirements.

¹⁸⁶ The clause 'and if you believe you will not F unless you intend to F' has been dropped for simplicity sake, but it is relatively easy to see how it can be included.

¹⁸⁷ As a side note, for those aware of the issue, contrary to Kolodny's objection in "Why Be Rational" the agent could revise his judgment about r2 on the basis of the elements of the r2 itself, or based on r1. The agent has ample basis to revise his judgment about r2.

Further, the suggested view will always generate rational requirements that are strict; i.e., any case in which an agent fails to take as a reason that which he recognizes to be one, the agent will be irrational. And, since there are two necessary and jointly sufficient conditions in the suggested view, it will always generate rational requirements that are broad; i.e., it will always be possible for an agent to remedy an irrationality by taking the reason to be the reason he recognizes it to be, or cease to recognize it as a reason.

By expanding the discussion of rational requirements in this section to not only include Broome-style rational requirements but also discuss Kolodny's critique of Broome as well as his own suggestion, it is hoped that the distinct character of the suggested view was cast in starker relief. Perhaps most distinctively the suggested view embraces aspects of each, disagrees with both, and argues that the debate between them centers on a question that is not deserving of the emphasis they both place on it.

Specifically it was argued that there is a fundamental rational requirement that concerns how agents reason. It was noted that this is evocative of, but different from, Kolodny's focus on process-requirements. It was also shown that the suggested view can plausibly generate Broome-style rational requirements that are both broad and strict. In the process, it also showed why Kolodny's arguments against wide-scope requirements are overly restrictive and should not be taken to be persuasive. Further, by showing this in Davidson's terms, it has also been shown that the suggested view is consistent with his descriptive account of akrasia.

Admittedly showing that the suggested view can generate Broome-style rational requirements that are broad and strict, reflects the intuitively appealing aspects of Kolodny's process-requirements and is also consistent with Davidson's descriptive account of akrasia is still far from showing that the suggested view points to a failure that is present in every case of irrationality, as is required by the Descriptive Requirement. Nonetheless, by showing that it is consistent with these leading candidates, the hope is that it has been shown to be at least as plausible as each of these alternatives. This point could be more firmly established by enumerating the canonical cases of irrationality in terms of the suggested view and pointing to the failure in each case. One could also generate an extensive list of Broome-style rational requirements based on the suggested view¹⁸⁸.

Having argued that the suggested view may plausibly satisfy the Descriptive Requirement, the discussion can turn its attention to the Normative Requirement, i.e., the failure to take as a reason that which one recognizes to be one must be of a special kind that can explain the normativity of irrationality. This is the subject to which the discussion will turn to next.

Normative Requirement

To begin it will be helpful to make a few comments about the normativity of rationality/irrationality.

¹⁸⁸ One caveat needs to be mentioned. The rational requirements one would generate would likely differ from Broome's own. To see this, simply consider Broome's first requirement, the one associated with having contradictory beliefs. Rather than seeing this as a failing of the suggested view, the promise of modifying Broome's rational requirements, which are based on insightful intuition, seems a virtue of a principled account of irrationality.

Consider the following example: Were Tom to knowingly put petrol in the car he believed to take diesel, given that he intended to put the right kind of fuel in the car, Tom's action would be irrational. This is merely a description of Tom's action, and bears no hallmark of normativity. As such, it is consistent with the account of irrationality so far described.

But claims about irrationality often carry the tenor of normativity, as would be the case if one were to admonish Tom by saying, 'Tom that's an irrational thing to do.' Or more explicitly, if we were to advise him by saying 'You ought to put diesel in the car.' Claims about (ir)rationality in these contexts have a normative character. The question is, what basis does the descriptive account of irrationality provide for understanding the source of normative claims of (ir)rationality?

From the outset, the account of rationality (or irrationality) that has been described is an account of proper (or improper) functioning. This might suggest that the normative criteria are already built in, as it were, to the descriptive account. For example, one can easily imagine the following normative claim consistent with this descriptive account: function properly.

But this does not capture the normative character of claims about rationality. Consider for instance the earlier example in which Tom was undertaking to put petrol in the car he believed took diesel. Those who would admonish Tom to instead put diesel in the car may well point to Tom's belief that it is a diesel car, but it seems very unlikely that they would admonish him to function properly. It also points to the ways in which the soldier case differs from that of the smoker.

Rather, the normative character of rationality claims stems from the normativity of reasons. Specifically, it stems from the normative force of the reason that the agent recognizes, but does not take to be a reason. This is in line with Scanlon's view and what Kolodny is suggesting in his Transparency Account.¹⁸⁹

To put the point more colorfully, when one admonishes Tom to put diesel in the car because he believes the car takes diesel, one is admonishing him to take as a reason that which he recognizes to be one. Or, one is admonishing him to do that which by his own lights he sees himself as having most reason to do.

In short, the suggested view offers the following answer to the Normative Requirement: the normative force of rationality/irrationality consists of the apparent normative force of reasons, nothing more.

Insofar as the suggested view explains how this normativity arises based on its answer to the Descriptive Requirement, the suggested view offers a plausible answer to the Normative Requirement, and thus to the Riddle of Irrationality.

Conclusion

This is a modest step towards the ambitious goal of offering an account of rationality. By focusing on the Riddle of Irrationality, it was possible to see the virtues and failings of accounts by Davidson, Scanlon and Broome. It was argued that the suggested view, which identifies irrationality with failing to take as a reason that which one recognizes to be one, meets the Descriptive and

¹⁸⁹ Kolodny, N. "Why Be Rational", *Mind*, New Series, Vol. 114, No. 455 (Jul., 2005) (pp. 509-563).

Normative Requirements that make up the Riddle of Irrationality, while sharing the virtues of Davidson's, Scanlon's and Broome's accounts. Specifically, it was shown that the suggested view is consistent with Davidson's account of akrasia, can generate Broome-style rational requirements, and provides a descriptive account that gives flesh to Scanlon's suggestion that the normativity of rationality stems from the reasons an agent recognizes.

Chapter 4: Preferences

The fact that there are multiple interpretations of 'preference' presents a challenge to Preference Theories; and undermines the justification of the completeness and transitivity conditions.

Further, seeking to address these challenges by exploring possible justifications of the completeness and/or transitivity conditions as conditions of rationality relies on a conception of rationality and a larger conception of practical reasoning.

As a result, the previous two chapters have begun to enumerate a conception of practical reasoning that has been emerging in philosophy of law, moral theory and philosophy of action. It is a conception of practical reasoning which takes reasons to be the basic normative concept. Reasons, on this view, are facts; which can be provided by value. Agents in turn, form judgments based on the reasons they recognize given their subsidiary beliefs; and rationality governs their reasoning.

With the backdrop established in previous chapters, it is now possible to begin to articulate a conception of preferences within this broader context of a conception of practical reasoning. This chapter will proceed in three stages. The first will stage will briefly discuss important characteristics of preferences that the suggested account will share with other accounts in the literature. The second stage will introduce the suggested account of preferences. And the third section will seek to face the challenges discussed to this point.

Stage 1: Characteristics of Preferences

While the account of preference suggested here is distinct from those discussed to this point, it shares commonalities with a number of them. The following discussion succinctly points to aspects of preferences the suggested account will seek to accommodate. It will also mention one account that has not been addressed to this point. Several should be made explicit. Specifically the following discussion will focus on the four concepts associated with preferences: choices, welfare judgments, value and all-things-considered judgments.

Choices

The connection between 'preference' and choice is perhaps the most fundamental. If Preference Theories are to guide and explain actions there must be a connection between preference and choices.

Indeed as was discussed earlier Revealed Preference Theory reduced preference to choices; as did the 'forcing procedure'. Though, both ran into difficulties because of this.

The suggested account will seek to maintain a strong connection between preference and choices but avoid the associated difficulties by resisting the temptation to reduce preferences to choices. Instead, it will take seriously a suggestion Sen made in criticizing Revealed Preference Theories, i.e., the

possibility of understanding 'preference' as the "attitudes or judgments that underlie choices."¹⁹⁰

Welfare judgments

As Sen writes "the normal use of the word permits the identification of preference with the concept of being better off."¹⁹¹ And the Money Pump argument relies on an interpretation of 'preference' as welfare judgments.

Yet there are those, like Mandler and Hausman who think that the association of 'preference' with welfare is a relic of the Utilitarian roots of Preference Theories whose time has passed. E.g., "Still, much confusion and pointless criticism would be avoided if locutions such as "welfare" or even "better off" were dropped."¹⁹²

While reducing 'preference' to welfare judgments is problematic – not least because of the tension it raises for the connection between 'preference' and choices discussed earlier – it is sometimes compelling to interpret an agent's preferences as reflecting his welfare judgments in certain circumstances. For example, it is natural enough to conclude that an agent's preference for saving money for retirement reflects his judgment that doing so will make him better off. The suggested view will seek to accommodate some connection between 'preference' and welfare judgments while also providing for the possibility that an agent's preferences may represent a broad range of other considerations.

¹⁹⁰ A.K. Sen, *Choice Welfare and Measurement*, (First Harvard University Press, 1998), Pages 54-73

¹⁹¹ (Sen 1977: 329)

¹⁹² Mandler, M. "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both", pg 8.

Value

In Weighing Goods: Equality, Uncertainty and Time¹⁹³, Broome offers an intriguing possibility for sidestepping the debate about welfare judgments. Broome suggests that the formalism of Preference Theories can be applied to value itself by substituting the preference relation with the betterness relation. As a result the alternatives suggested by the formalism are the better alternatives, because they reflect the betterness relation itself. But Broome establishes the connection between value and the formalism of Preference Theories by giving up the preference relation. The subject of consideration is no longer an agent's preferences, but value.

Broome's suggestion puts value at the center of questions about what ought to be done in a way that is rare in Preference Theories, and the suggested view will seek to maintain the focus on this relationship. However, it will not follow Broome's example of replacing the preference relation with the betterness relation for two reasons. First, the current effort is concerned with offering an account of preferences, and so following Broome would be contrary its aims. And perhaps more substantively, there is good reason to think there is a plurality of value; and if this is the case it is unlikely that value can be cashed out in terms of a betterness relation that satisfies the axioms of Preference Theories.

All things considered

¹⁹³ Broome, J, Weighing Goods: Equality, Uncertainty and Time, Oxford : Cambridge, Mass. : Basil Blackwell, 1991.

In his forthcoming manuscript, Hausman¹⁹⁴ recognizes the challenges that have plagued Preference Theories as the result of the multiple interpretations of 'preference'. However, contrary to Sen, he suggests that rather than think of 'preference' as admitting multiple interpretations, economists should legislate a canonical understanding of preferences as total comparative evaluations. Such total comparative evaluations may reflect an agent's welfare judgments, but they may also reflect an agent's evaluation of the alternatives on any of a number of different grounds such as relative value. Such total comparative evaluations, or all-things-considered judgments as they will be referred to here, are evaluations based on all the considerations the agent takes to be relevant.

The ability of Hausman's suggestion to accommodate both the common association of preferences with welfare judgments and the connection between preferences and value count in its favor. If connected to choice through an appropriate conception of practical reasoning, Hausman's suggestion presents a promising account of preference.

It is on this last point that there is the greatest difference between Hausman's account and the current suggested view of 'preference'. The suggested view will follow Hausman's suggestions for the interpretation of 'preference', but locate it within the context of the conception of practical reasoning developed in earlier chapters rather than the belief-desire view enumerated by Hausman.

Here is why.

¹⁹⁴ Hausman, D., Preference, Value, Choice, and Welfare, Cambridge University Press, NY, NY, 2012.

The conception of practical reasoning Hausman relies upon is sketched briefly. Nonetheless, it is possible to infer what Hausman may have in mind from the few comments he does make and the broader literature.

In discussing the relevant conception of practical reasoning, Hausman is predominantly concerned with explaining actions. He notes two different perspectives from which an action can be explained, the first and third person perspective. From the first person perspective, “The question agents ask themselves is not “Given my beliefs and desires, what do I predict that I will do?” but “What *should* I do” or “What do I have most *reason* to do?””¹⁹⁵ An agent’s beliefs and desires may not be decisive, because the agent is aware that his beliefs may be faulty and his desires are open to revision based on his understanding of the relevant reasons. From the agent’s perspective it is the facts and value that motivate his action.

Yet from the third person perspective psychologists and philosophers explain the agent’s action based on a “belief-desire psychology”¹⁹⁶. The basic idea is that “Beliefs are linked to reasons, because beliefs purport to provide agents with facts, and facts can be reasons.”¹⁹⁷ and “desire presents its object as in some regard valuable or “to-be-done” .¹⁹⁸

The belief-desire psychology Hausman presents may be a conception of practical reasoning that is consistent with Preference Theories. Chapter 1:

¹⁹⁵ Hausman, D., Preference, Value, Choice, and Welfare, Cambridge University Press, NY, NY, 2012, pg 5.

¹⁹⁶ Hausman, D., Preference, Value, Choice, and Welfare, Cambridge University Press, NY, NY, 2012, pg 5.

¹⁹⁷ Hausman, D., Preference, Value, Choice, and Welfare, Cambridge University Press, NY, NY, 2012, pg 6.

¹⁹⁸ Hausman, D., Preference, Value, Choice, and Welfare, Cambridge University Press, NY, NY, 2012, pg 6.

Introduction acknowledged that there may be many such views of practical reasoning. However, one difficulty with this view, and the reason it will not be pursued in the current effort, is that it precludes asking perhaps the most compelling question by omitting normative reasons.

People occupy a normatively complex world and employ their judgment in order to navigate. A fruitful conception of rationality and practical reasoning should seek to explain, predict *and* guide people's choices in light of this. By omitting the concept of normative reasons, Hausman's account does not have a meaningful way to inquire about what agents should do. To be clear, the point is not that Hausman omits reference to normative reasons in his account of preferences. Since his goal is largely explanatory and predictive this would be an unfair criticism to level against him. Rather the point is that on the belief-desire psychology Hausman describes, the possible answer to normative questions would merely be based on value (or the believed value) of the outcome constrained by beliefs. This seems to substantially reduce the normative to the evaluative; and there is far more to the normative than just this.¹⁹⁹

As a result, while the suggested view will follow Hausman's lead in interpreting 'preference' as all things considered judgments²⁰⁰, it will not adopt his belief-

¹⁹⁹ For further elaboration on this see the description of the types of reasons and the ways in which they compete in Chapter 2: Reasons.

²⁰⁰ Minor differences remain between the Hausman's interpretation of 'preference' and that of the suggested view. For example, Hausman takes preferences to be inherently comparative. The suggested view can accommodate the possibility of an agent simply preferring an alternative, rather than necessarily preferring an alternative over another. Further, Hausman would likely object to the characterization of 'preference' as judgments, and rather refers to them as evaluations. And for Hausman beliefs play a distinct role from preferences, whereas on the suggested view preferences reflect an agent's beliefs. As a result, it may be more fruitful to associate the suggested view of 'preference' with Hausman's concept of 'final preference'.

desire psychology. Rather, it will attempt to situate the concept 'preference' within the conception of practical reasoning developed to this point.

Characteristics of Preferences

Taken together this discussion presents four characteristics of preferences that the suggested view will seek incorporate:

- a fundamental connection between preference and choice
- a connection between preferences and welfare judgments
- a connection between preferences and evaluation of value
- preferences as a reflection of an agent's all things considered judgments.

Stage 2: Suggested Account of Preferences

Recall that one of the virtues of Davidson's account of irrationality discussed in the previous chapter was that it offered a description of the extenuated relationship between an agent's judgments and his choices. It could accommodate the fact that an agent may arrive at a number of different judgments conditional on relevant considerations, yet arrive at an unconditional judgment about how to choose. Davidson identifies the second of these judgments with intentions. The following section will explore the possibility of identifying the first of these judgments with preferences.

To do so, it may be helpful to have a brief restatement of the key elements in Davidson's account:

"In deliberating about whether to S, an agent considers her reasons for S-ing. She may believe she has reasons for and against S, though the conditional

form of the agent's attitudes, Davidson suggests, ensures that even in such cases the agent will not thereby be entertaining a contradiction. Following Davidson, these attitudes can be called prima facie judgments (PF-J's), and they can be represented as follows: pf(evaluative judgment; evidential grounds) or as pf(action type; reasons).

An agent may simultaneously arrive at conflicting PF-J's such as pf(s, r_1) and pf(not s, r_2). In such cases, the agent resolves the conflicts by arriving at an all things considered judgment (ATC-J), which is a single conditional attitude based on all the relevant reasons, i.e., pf(s, r_1 and r_2).

In contrast, intentions, the attitudes on which an agent acts, are unconditional in form. Such intentions, or sans-phrase judgments (SP-J's), can be represented as follows: sp(s).²⁰¹

Further it was argued, that in cases of rational choice an agent chooses, and forms his SP-J's, based on the reason which ground his all-things-considered-judgments.²⁰²

With this brief reminder of Davidson's account, it is possible to cache out the suggestion more explicitly. Whereas Davidson identifies SP-J's with intentions, the current suggestion is to identify ATC-J's with preferences. Note, since an agent's ATC-J is a PF-J that is based on all the relevant considerations, at times preferences will be discussed in terms of PF-J's.

²⁰¹ From Chapter 3: Rationality.

²⁰² Recall the discussion of akrasia in Chapter 2: Rationality.

With this brief reminder of Davidson's account, it is possible to cache out the suggestion more explicitly. Whereas Davidson identifies SP-J's with intentions, the current suggestion is to identify ATC-J's with preferences. Note, since an agent's ATC-J is a PF-J that is based on all the relevant considerations, at times preferences will be discussed in terms of PF-J's.

On first glance, this may be a surprising suggestion, for it may not be obvious why this should be an account of preferences. A few reflections should help allay those concerns. Let's take a simple case, for example an agent's preference for ice cream over fruit for dessert. Using the current vocabulary one can describe this agent as having arrived at a PF-J for ice cream over fruit for dessert. The agent may have arrived at this PF-J based on the fact that the agent likes the taste of ice cream better than he likes the taste of fruit, which he recognized to be a reason.

Further, given that there are no other reasons that the agent takes to be relevant, the agent's PF-J reflects all the relevant considerations, and as such is the agent's ATC-J. Were the agent ordering dessert, he may well form the rational intention to have ice cream, rather than fruit, for dessert.

Imagine the agent in this circumstance becomes aware of new information. The agent comes to learn that it is healthier to have fruit than ice cream for dessert. The agent recognizes that this fact is a reason for having fruit, rather than ice cream, for dessert. And in so doing he recognizes it as a relevant consideration. He forms a conditional judgment, a PF-J, for fruit over ice cream for dessert.

Whereas the agent had previously arrived at an ATC-J, he now has two distinct PF-Js. In this circumstance, the agent does not need to revise either judgment. They do not conflict. He has one PF-J for ice cream over fruit for dessert based on taste; and another PF-J for fruit over ice cream for dessert based on health considerations. Both of the agent's PF-J's are conditional.

Note that the same would not be true if the judgments in question were intentions rather than preferences. For example, if the agent intended to have ice cream, rather than fruit, for dessert, and intended to have fruit, rather than ice cream, for dessert, then the agent would have had conflicting judgments. This is a reflection of the unconditional nature of intentions; and points to one of the ways in which intentions and preferences differ.

However, were the agent in question to consider what to order for dessert, rationality would require the judgment he arrived at to reflect both of the reasons he understood to be relevant. That is to say it may be rational for the agent to arrive at a preference for ice cream over fruit for dessert, or for fruit over ice cream for dessert, but rationality would require whichever judgment he arrived at to reflect both of the reasons he recognized. And this is a result of the fundamental requirement of rationality discussed in the previous chapter.

If this discussion helps shed light on the conditional nature of preferences, it also provides the basis for addressing two related concerns about the suggested view. The first is that preferences, understood as ATC-J's, do not seem to share much in common with the desires often associated with

preferences. Hopefully the example just discussed illustrates why these concerns can be put to rest. There are times when one's preferences reflect one's desires, tastes, likes, etc. Preferences between vanilla and chocolate ice cream may be a ready example. And many of one's preferences may be formed on the basis of such considerations. However, as one becomes aware of other considerations, one's preferences may change. As in the earlier example, the agent may like the taste of ice cream more than the taste of fruit yet form a preference for fruit over ice cream because of health considerations. Indeed there may be a wide range of such considerations that come to form the basis of an agent's judgments. It is not that the suggested view does not give a significant role to such things as desires, tastes and the like, but that the current account suggests that they have a place within a much larger normative context; and this is reflected in the basis for agents preferences.

A related concern may be that ATC-J's are not sufficiently different from beliefs to be distinguished from them. Does the suggested view merely reduce preferences to conditional beliefs? It is easy to see why this concern is not warranted either. Recall the discussion in Chapter 3: Rationality about the distinction between recognizing a reason and believing that there is a reason. In that discussion it was shown that the two are strictly separable. Similarly in this context, the differences between the two comes to the fore. To see this, consider the earlier example, but substitute beliefs about conditional relationship for PF-J's. To make the example explicit, the agent has a belief that for health reasons it is better to choose fruit over ice cream for dessert. And the agent has a belief that for taste reasons it is better to choose ice cream over fruit for dessert. When choosing what to order for dessert, it would be

sensible for the agent to do so on the basis of both of these beliefs, but there is no rational requirement for him to do so precisely because he recognizes neither as reasons for his choice. It is in part the conditional nature of preferences as all things considered judgments that distinguish them from beliefs, and makes them uniquely suited to play this distinctive role.²⁰³

Further, the suggested view has several of the attractive features of an account of preference discussed earlier. For example it offers a way of understanding Sen's suggestion: ATC-J's are the judgments which underlie choices, and by extension, preferences understood as ATC-J's are the judgments which underlie choices. Further, on this reading there is an intimate connection between preferences and choices, yet preferences do not reduce to choices. Rather, the connection between preferences and choices is rationality governed. For example, recall that on Davidson's account if an agent chooses contrary to his ATC-J, he makes an irrational choice, i.e., his choice is akratic. Described in terms of preferences, this can be restated to say that if an agent chooses contrary to his preferences, his choice is irrational.

In addition, on this suggestion, preferences reflect an agent's evaluation, without being narrowly reduced to welfare judgments. An agent forms an ATC-J based on an evaluation of all the considerations he takes to be relevant. The

²⁰³ Note that for those still reluctant to understand preferences as ATC-J's, it is possible to suggest that preferences are somehow a distinct mental state from ATC-J's, but rationally related to them. Nothing in the argument presented here precludes that possibility. After all, as discussed at length in the introduction, the focus of the current effort is to enumerate a potential account of preference situated against a rich backdrop of a conception of practical reasoning; rather than argue against the possibility of other plausible accounts. That said, it should be noted that alternative plausible accounts are not yet readily available. And, further, if one sought to distinguish preferences from ATC-J's in the way described, it would seem incumbent to argue that this new ontological category is significantly distinct from ATC-J's to warrant independent consideration. And, if the new suggestion shares the same advantages as the current account, explain how it does so. While this may be entirely plausible, at present it is difficult to be optimistic about the prospects for such an approach.

agent's ATC-J reflects his evaluation of the alternatives he is confronted with. As a result, identifying an agent's preferences with his ATC-J, the agent's preferences reflect his evaluation of the alternatives. So the suggested view is able to account for the fact that an agent's preferences reflect his evaluation of the alternative, without thereby reducing preferences to welfare judgments narrowly prescribed.

If these considerations recommend the suggested view, the pressing question is whether the suggested view can shed light on the justification of the completeness condition, offer a justification of the transitivity condition and address the possibility of fine individuation.

Stage 3: Facing the Challenges

With this brief statement of the suggested account, it is possible to explore the challenges raised to this point. Specifically this stage will be comprised of three sections. The first will explore the possible justification for the completeness condition. The second will seek to address the challenge posed by the prospect of fine individuation. And the third will suggest a possible justification for the transitivity condition as a condition of rationality.

Completeness Condition

The earlier discussion of the completeness condition in Chapter 1: Introduction questioned the justification for the completeness condition as a condition of rationality for two reasons. First, the standard argument for justifying the completeness condition, i.e., the 'forcing procedure', relies on a different interpretation of 'preference' than the associated conception of rationality. And

second, the completeness condition itself is too demanding for a condition of rationality. Both will be looked at briefly before suggesting a different way of thinking about the completeness condition.

Recall that the 'forcing procedure' relied on an interpretation of 'preference' as choice. It suggests that any agent whose preferences are possibly incomplete can be forced to choose between two alternatives by putting them in a situation in which he will end up with a less preferred third alternative if he does not make a choice. Since the third alternative can be made sufficiently unattractive, it is suggested that a choice will always result; and this choice can be interpreted as the agent's preference between the two alternatives.

The difficulty is that the 'forcing procedure' is a poor fit for the associated consequentialist conception of rationality. The basic idea was that features of an agent's preference could be shown to be irrational, if those features led the agent to suffer harm. If the agent failed to make a choice when faced with the 'forcing procedure' his preferences would be incomplete and he would suffer harm as a result. Therefore having incomplete preferences is irrational.

The issue with this is that in order to establish that the agent in question suffered harm, this justification would need to rely on an interpretation of 'preference' as welfare judgments. Since the 'forcing procedure' relied on an interpretation of 'preference' as choice, it cannot establish that the agent in question suffered harm. And as a result, the 'forcing procedure' does not justify the completeness condition as a condition of rationality.

Further, it was suggested that there is independent reason to think that this is not necessarily a failure of the ‘forcing procedure’ or the associated conception of rationality. The completeness condition, if taken literally, is extremely demanding. There are an infinite number of possible alternatives. To suggest that rationality requires an agent to form a preference between all possible pairs of alternatives is excessive and possibly a reductio of the associated conception of rationality. For an agent can rationally fail to form a preference between two alternatives by merely failing to be aware of them as alternatives.²⁰⁴

The suggested view will fare little better in offering a justification for the completeness condition as a condition of rationality. And the reason for this is straightforward. Recall the discussion in Chapter 3: Rationality of the case in which the agent fails to recognize a consideration as a reason, and therefore does not take it to be a reason. It was argued that this was not a case in which the agent was irrational, though it may include cases in which the agent was unreasonable. If the agent rationally fails to recognize a reason that pertains to an alternative, and therefore fails to form a preference related to the alternative, it would seem there is nothing in the suggested view that could justify considering the agent irrational.

Based on this it was suggested that it may be more fruitful to think of the completeness condition as an idealizing assumption; i.e., an assumption that is not literally true, but the stipulation of which makes the subject matter formally

²⁰⁴ This is only one way in which an agent’s preferences may be rationally incomplete. There are others. For example, it is possible that there are hard choices. i.e., choices in which reasons do not answer the question, what ought to be done? There may well be circumstances in which the agent has sufficient reason to do x, and sufficient reason to do y, and yet there is no conclusion about what one ought to do. “Sophie’s Choice” is one likely example. The question is subject to serious debate. For a detailed discussion of this topic see Levi, I, Hard Choices: Decision making under unresolved conflict, Cambridge ; New York : Cambridge University Press, 1986.

tractable. Specifically, in “Unreal Assumptions in Economic Theory”²⁰⁵, Musgrave surfaces an interesting possibility.²⁰⁶ He notes that different types of assumptions play different roles in economic theories²⁰⁷. For example, some assumptions point to factors, which might be expected to have an effect on the subject of the theory, and claim that they will have negligible effect. Musgrave refers to assumption of this sort as “negligibility assumptions”.²⁰⁸ The completeness condition is not a negligibility assumption. For if it turned out that agents did not have complete preferences, preferences would not easily admit to the orderings which are the currency of much of Preference Theory.

In contrast to negligibility assumptions, Musgrave also introduced the notion of domain assumptions. Domain assumptions “specify the *domain of applicability*”²⁰⁹ of a theory by pointing to factors that must be present, or absent, for a theory to be applied. The completeness condition has the hallmarks of a domain assumption. That is to say that Preference Theories can

²⁰⁵ Musgrave, A., “‘Unreal Assumptions’ in Economic Theory: The F-Twist Untwisted”, *Kyklos* 34: 377-387 August 1981.

²⁰⁶ There are of course other attempts to justify the completeness conditions. For example in his forthcoming book, Hausman presents an intriguingly concise explanation for the relationship between rationality and the completeness condition. “Completeness is a boundary condition on rational choice. An inability to compare alternatives is not itself a failure of rationality, but when people are unable to compare alternatives, they are unable to make a choice on the basis of reasons.” (Hausman 2012, pg19)

Unfortunately it is unsuccessful for two reasons. First, the claim that agents who are unable to compare alternatives are unable to make a choice on the basis of reasons is false. The mere fact that an agent is unable to compare x and y does not mean that he does not recognize that there are reasons for x and reasons for y. He can choose x based on the reasons for x. And he can similarly choose y based on the reasons for y.

Second, it appears to rely on a conflation of completeness and comparability, which are strictly different. There are many reasons an agent may have incomplete preferences. For example, an agent may simply not be aware of the available alternatives. Or the alternatives may be incommensurable yet comparable. An agent’s inability to compare alternatives may be only one reason why an agent would have incomplete preference.

As a result, Musgrave’s suggestion seems more promising.

²⁰⁷ Presumably Musgrave’s insights apply more broadly to theories in other domains as well.

²⁰⁸ Musgrave, A., “‘Unreal Assumptions’ in Economic Theory: The F-Twist Untwisted”, *Kyklos* 34: 377-387 August 1981, pg 378.

²⁰⁹ Musgrave, A., “‘Unreal Assumptions’ in Economic Theory: The F-Twist Untwisted”, *Kyklos* 34: 377-387 August 1981, pg 381.

be applied in circumstance in which agents have, or may have, complete preferences.²¹⁰

As a domain assumption the justification for the completeness condition no longer relies on the associated conception of practical reasoning in the same way. Rather the justification of a domain assumption rests on its ability to demarcate those domains in which the theory may apply. Since the completeness of agents preferences are implied by the fact that the agent's preferences admit to an ordering, the completeness condition is one of the ways to demarcate a domain in which an agent's preferences admit to an ordering and become formally tractable within Preference Theories. If this is true, it suffices as a justification for the completeness condition as a domain assumption.

Here the role of the associated conception of practical reasoning is quite different. Rather than justifying the completeness condition, it sheds light on what will be required to satisfy the condition. In this sense the suggested view has an advantage over views such as MHV. For the suggested view offers an account of how agents form preferences based on recognized reasons. For example on the suggested view an agent can rationally fail to form a preference between two alternatives, if he is unaware of them as alternatives. As a result, satisfying the completeness condition will likely require that agents have full knowledge of, or at least awareness of, the available alternatives. Further, since an agent may also rationally fail to form a preference between two

²¹⁰ This is admittedly a significant claim. The current trend is to apply Preference Theories ever more broadly. And, if the completeness condition is a domain assumption, this would suggest that some current applications go beyond the domains to which the theory can be fruitfully applied.

alternatives if they rest on incommensurable or incomparable values, the completeness condition may obtain in circumstances in which the agent recognizes reasons based on a single value. It is not coincidental that version of these requirements are often associated with the application of Preference Theories. And the fact that the suggested view can shed light on these aspects of a completeness condition justified as a domain assumption should count in its favor.

Fine Individuation

In two influential works²¹¹, Broome showed that the prospect of fine individuation poses a challenge to the transitivity condition. For if any apparent instance of intransitivity could be explained away as a case in which the agent had transitive preferences over more finely individuated alternatives, then transitivity would not meaningfully constrain an agent's preferences. The transitivity condition would be vacuous.

The claim to fine individuation rested on the conjunction of two further claims. First, that while two alternatives appear to be identical, they are in fact different alternatives. And second, that it is rational to have preferences between these alternatives.²¹²

²¹¹ Broome, John, "Can a Humean be moderate?", in: Frey, R G (ed), *Value, Welfare, and Morality*, Cambridge University Press : New York, 1993. And, Broome, John, *Weighing Goods*, Oxford : Cambridge, Mass. : Basil Blackwell, 1991.

²¹² Broome introduced to Principle of Individuation by Justifiers and the Rational Requirements of Indifference to address these respective claims and address the challenges presented by the prospect of fine individuation.

The suggested view has the resources for addressing the challenge posed by the prospect of fine individuation by offering a principled means for denying claims of the first kind in cases of irrationality.²¹³

The basic idea can be expressed in the Normative Identity Claim (NIC), which is the conjunction of two further claims.

- Claim 1: if two alternatives are identical, or are to be treated as identical, then they must be normatively indistinguishable. Or equivalently: if on the other hand, there is a normatively salient difference between two alternatives, then they should not be treated as identical.²¹⁴
- Claim 2: if two alternatives are normatively indistinguishable, then they will be recommended by the same relevant reasons with the same force; and if two alternatives are not recommended by the same relevant reasons with the same force, then they are not normatively indistinguishable.²¹⁵

As a result of NIC, one can establish the identity of alternatives by establishing the identity of relevant reasons. This yields the following variation of Broome's Principle of Individuation by Justifiers, PIJ': alternatives should be individuated from one another, if and only if they differ in terms of relevant recognized reasons.

There are a couple of things about this that are worthwhile to note. First, the current discussion is interested in how PIJ' applies to alternatives as the agent

²¹³ This is essentially the role of Broome's Principle of Individuation by Justifiers (PIJ) discussed earlier.

²¹⁴ This will be taken to be analytic.

²¹⁵ This is a result of the status of reasons as the basic normative concept.

understands them, not necessarily to the actual alternatives²¹⁶. The reason for this is simple. PIJ' applies to either actual alternatives or to the alternatives as the agent understands them. If PIJ' is based on the actual reasons, it will serve as an identity condition for the actual alternatives. If on the other hand it is based on the recognized reasons, it will serve as an identity condition for the alternatives as the agent understands them. Since the current discussion is interested in the transitivity condition as a condition of rationality, it is interested in the agent's reasoning based on recognized reasons²¹⁷.

The second thing to note is that the identity conditions as described are recursive. In other words, establishing the identity of alternatives depends on establishing the identity of recognized reasons. And, given the discussion of the Complete Reasons Identity Relation (CRIR)²¹⁸ in Chapter 2: Reasons, establishing the identity of recognized reasons depends on establishing the identity of the agent's beliefs about the elements of complete reasons. As a result, on this account preferences are deeply intensional, not only in their issue, but in their object.

²¹⁶ NB: PIJ' applies to both alternatives as the agent understands them, and alternatives as they are.

²¹⁷ In case there is any doubt about this, simply consider the two relevant counter-examples. E.g., the agent's preferences are apparently intransitive. In the first case it is possible to establish the identity of the alternatives as the agent understands them, but the actual alternatives are distinct; and the case in which the alternatives as the agent understands them are distinct, while it is possible to establish the identity of the actual alternatives.

In the second case, the agent would be guilty of two errors. First, he is guilty of having intransitive preferences. And second, he is guilty of making a mistake about the nature of the alternatives. In the first case, the agent would not be guilty of having intransitive preferences though he would be guilty of making a mistake about the nature of the alternatives.

²¹⁸ Complete Reason Identity Relation (CRIR). CRIR: the complete reason $R''(a_1, o_{v_1}, x_1)$ is identical with the complete reason $R''(a_2, o_{v_2}, x_2)$ if and only if $a_1 \equiv a_2$, $o_{v_1} \equiv o_{v_2}$, and $x_1 \equiv x_2$. Where $R''(a, o_v, x)$ can be understood as: agent's, x 's, taking an action, a , brings about an outcome or increases the likelihood of, o , which realizes or preserves some value, v .

One consequence of this is that establishing that an agent has intransitive preferences can be a difficult affair. It depends not just on the choices he makes, but on how he understands the alternatives he is choosing over, the considerations that he views as relevant and his beliefs about the facts of the matter. Rather than a drawback of the suggested account, this seems a virtue. For convicting an agent of irrationality is a high charge, one that may be leveled too quickly against an agent who may instead simply have different beliefs about what is at stake in a choice situation. Without an articulated account of how agents choose, too many faults may get wrapped under the banner of irrationality.

Of course, in many situations it is not necessary to go to such lengths to understand an agent's preferences. Much of the information may be supplied by common knowledge or a shared context.

This also suggests another possibility that makes too little appearance in the literature. Given that people are generally rational and that preferences reflect an agent's views about the considerations that are relevant to choice situations, cases of apparent intransitivity may convey important information about the choice situations themselves. This is information which may be too quickly discarded if the agents are hastily convicted of irrationality.²¹⁹

Returning to the question of fine individuation, this discussion shows that while the question of how to individuate alternatives in specific choice situations presents challenges, the suggested account offers a principled means for doing so that reflects the nuances of how agent's form preferences. As a result,

²¹⁹ The Sure-Thing Principle and its associated paradoxes seems a quintessential example.

further individuation is not always possible, and there is no threat of rendering the transitivity condition vacuous. In this way the suggested account can address the challenges posed by the prospect of fine individuation and defend a meaningfully transitivity condition.

Justifying Transitivity:

The transitivity condition is widely regarded to be a condition of rationality²²⁰, yet it lacks a compelling justification. For example the discussion in Chapter 1: Introduction showed that the Money Pump, which is often relied on to justify the transitivity condition is more effective as a poignant example than a justification.

This section will explore the possibility of offering a justification of the transitivity condition against the backdrop of the suggested conception of practical reasoning. The discussion will begin by examining a case of intransitivity in a stipulated idealized situation. Then it will render a case of intransitivity in-line with the suggested account of preferences and seek to locate the rational failure associated with intransitivity. Relying on the account of rationality offered in the previous chapter, it will claim that failures of this type constitute an irrationality, and as such suggest that the transitivity condition is a justified condition of rationality in idealized circumstances. It will then relax the idealizing assumptions and explore whether this holds robustly in different circumstances.

²²⁰ For example: "The details of decision theory are not universally agreed. Different versions have different axioms. But they do all agree at least on the axiom of transitivity. Transitivity is a minimal condition of consistency: if consistency does not require transitivity, it requires nothing (footnote omitted). So we may take it that all moderate Humeans believe rationality requires a person to have transitive preferences." Broome, John, "Can a Humean be moderate?", Ethics Out of Economics, Cambridge; New York : Cambridge University Press, 1999, pg 70.

Intransitive preferences – The generic case

Consider the generic case of violations of the transitivity condition in which the agent is presented with three sequential choices between x and z , z and y , and y and x , respectively. As before, the agent chooses z over x , y over z , and x over y . The agent's preferences can be represented as violations of the transitivity condition: $x > y$, $y > z$, $z > x$.

As before, incorporating the challenge posed by fine individuation, this can be restated as follows: $x_3 > y_3$, $y_2 > z_2$, and $z_1 > x_1$.

Now consider the associated judgments. The agent forms a preference, that is to say an all-things-considered judgment, for x over y based on his conditional judgments for x and his conditional judgments for y . These can be represented as $pf(x_3, r_{x3})$ and $pf(y_3, r_{y3})$ ²²¹. And the associated all-things-considered judgment can be represented as $atc(x_3, r_{x3}$ and $r_{y3})$. The agent's other preferences admit to similar representation yielding the following statements of the relevant judgments:

- Choice 1: $atc(z_1, r_{z1}$ and $r_{x1})$ based on $pf(z_1, r_{z1})$ and $pf(x_1, r_{x1})$
- Choice 2: $atc(y_2, r_{y2}$ and $r_{z2})$ based on $pf(y_2, r_{y2})$ and $pf(z_2, r_{z2})$
- Choice 3: $atc(x_3, r_{x3}$ and $r_{y3})$ based on $pf(x_3, r_{x3})$ and $pf(y_3, r_{y3})$

As before, this would be a violation of transitivity if: $x_1 \equiv x_3$, $y_2 \equiv y_3$, and $z_1 \equiv z_2$.

And by PIJ', $x_1 \equiv x_3$, $y_2 \equiv y_3$, and $z_1 \equiv z_2$, if $r_{x1} \equiv r_{x3}$, $r_{y2} \equiv r_{y3}$, and $r_{z1} \equiv r_{z2}$.

²²¹ The formalism has been augmented to reflect the fact that the reasons at hand are reasons that are relevant to the alternatives in question. The possibility of indexing the alternative to the choice situation, which posed the problem of fine individuation, has also been maintained so as not to presume that all the associated challenges have been addressed.

Since the goal is to explore possible justifications of the transitivity condition by considering violations of it, assume $r_{x1} \equiv r_{x3}$, $r_{y2} \equiv r_{y3}$, and $r_{z1} \equiv r_{z2}$, and as a result $x_1 \equiv x_3$, $y_2 \equiv y_3$, and $z_1 \equiv z_2$, which would yield the intransitive preferences: $x > y$, $y > z$, $z > x$ ²²².

Idealizing assumptions

On the suggested view, the rationality of preferences depends on the reasons an agent recognizes. As a result, whether it is rational for an agent to have intransitive preferences will depend on the reasons an agent recognizes and judgments he forms on that basis.

As was mentioned in Chapter 2: Reasons, there are a multitude of reasons and they compete in a variety of ways. There are first order reasons and second order reasons. There are reasons to perform an action for its own right and reasons to perform an action because of the consequences it will bring about. There may even be reasons based on deontological considerations. While it was acknowledged that the issue of how reasons compete is deserving of considerable attention, to make the issue tractable for the moment assume the reasons at hand are consequentialist first order reasons to take an action based on the valuable outcome that will come about.

Even so, it is possible that there is a plurality of value in the world. And if there is, reasons may also compete based on the kind, as opposed to merely the extent, of value that the outcome will realize. Once again, to make the issue tractable, assume that there is a single kind of value that is relevant to the choice situation considered.

²²² Relying on the assumed identity, the subscripts have been omitted.

Justifying the transitivity condition

Consequentialist first order reasons for an agent to take an action can still compete in more than one way. Two different outcomes may realize one and the same value to a greater or lesser extent. And the agent may regard two different actions as having a higher or lower likelihood²²³ of realizing the valuable outcome.

In considering whether to x or y, the agent considers which alternative is more likely to bring about the more valuable outcome.²²⁴ Similarly in considering whether to y or z, the agent considers which alternative is more likely to bring about the more valuable outcome. And in considering whether to z or x, the agent considers which alternative is more likely to bring about the more valuable outcome.

As before, these judgments can be rendered to make explicit the relationship between the agent's judgments regarding an alternative and the grounds they are based on. For the sake of simplicity, focus on the agent's judgments of the individual alternatives, i.e.,

- x: $pf_1(x_1, r_{x1})$ and $pf_3(x_3, r_{x3})$
- y: $pf_2(y_2, r_{y2})$ and $pf_3(y_3, r_{y3})$
- z: $pf_1(z_1, r_{z1})$ and $pf_2(z_2, r_{z2})$

²²³ Here it is the agent's view of how likely an action is realize the value of the outcome that matters, as opposed to how likely it is that an action realize the value of the outcome, because it is the agent's recognized reasons that are at issue.

²²⁴ For present purposes the question of how the likelihood and the value of the outcome combine can be left up to the agent. What is important is that the agent arrives at a way judgment based on the reasons he recognizes.

Since this is a case in which PIJ' and CRIR hold, this can be re-stated as follows:

- x: $pf_1(x, r_x)$ and $pf_3(x, r_x)$
- y: $pf_2(y, r_y)$ and $pf_3(y, r_y)$
- z: $pf_1(z, r_z)$ and $pf_2(z, r_z)$

Either the agent arrived at the same PF-J when considering the same alternative in each choice situation²²⁵ or he did not²²⁶. If he deliberated based on the same reasons, but arrived at different judgments, then the reasons on which he deliberated played a different role in his deliberations. Even if the agent gave the recognized reason the appropriate role in deliberation in one instance, he did not give it the appropriate role in deliberation in both instances. This is a violation of the fundamental requirement of rationality enumerated in the previous chapter, and the agent is guilty of an irrationality.²²⁷

Further if $z > x$, $x > y$, and $y > z$, then it cannot be the case the agent arrived at the same judgments about the same alternatives based on the same reasons. I.e., it cannot be the case that:

- x: $pf_1(x, r_x) \equiv pf_3(x, r_x)$
- y: $pf_2(y, r_y) \equiv pf_3(y, r_y)$

²²⁵ For example, this can be represented as follows:

- x: $pf_1(x, r_x) \equiv pf_3(x, r_x)$
- y: $pf_2(y, r_y) \equiv pf_3(y, r_y)$
- z: $pf_1(z, r_z) \equiv pf_2(z, r_z)$

²²⁶ For example, this can be represented as follows:

- x: $pf_1(x, r_x) \equiv pf_3(x, r_x)$
- y: $pf_2(y, r_y) \equiv pf_3(y, r_y)$
- z: $pf_1(z, r_z) \neq pf_2(z, r_z)$

²²⁷ Of course it is possible for the agent to rationally form different judgments about x in the subsequent choice situation, if for example his preference was based on a different reason. But, by AIC, in that case the agent would be considering a different alternative, and so would his preferences would not be intransitive.

- $z: pf_1(z, r_z) \equiv pf_2(z, r_z)$

So, if the agent's preferences were intransitive, then the agent's ATC-J's must have tracked at least one instance of an irrational PF-J. I.e., at least one of the agent's ATC-J's must have been irrational. In this instance, given the consequentialist first order reasons to realize a single kind of value, having intransitive preferences would be irrational. In this instance, given the idealizing assumptions, the suggested view can justify the transitivity condition.

It appears the suggested view satisfies a key desiderata of an account of preferences. However, to leap to this conclusion may be too quick. For only a narrow subset of reasons were considered. What happens if the restrictive conditions are relaxed?

Relaxing the assumptions – a plurality of values

Given the multitude of different reasons and ways in which reasons can compete, it is not feasible to walk through each dimension in which the conditions can be relaxed. Nonetheless, it is informative to look at one possibility. What happens if there is a plurality of value?

If there is a plurality of value, this justification of the transitivity condition does not hold. The reason for this is simple. Consider the following three choice situations. Michael is presented with the choice between staying home (h) and going out with friends (f), then the choice between going out with friends and going to the museum (m), and finally the choice between going to the museum and staying home. In each case Michael chooses the first option. His reasons

are as follows. In the first instance Michael chooses to go out with friends, because it has been a long-time since he has done so and he would like to maintain the friendships he values. In the second instance, Michael chooses to go to the museum because it has a rare display of painting that would be valuable to engage with. And in the third instance, he chooses to stay home, because he has a regular routine with his family that he cherishes and would like to perpetuate. Michael's preferences can be represented as follows: $h > f$, $f > m$, and $m > h$. Michael's preferences in this example are intransitive.²²⁸ Yet Michael's preferences may be rational all the same. For if there is a plurality of value it is possible that in each instance Michael chose the more valuable alternative and the alternative he had more reason to choose. If there is a plurality of values, value itself may be intransitive²²⁹. And the balance of

²²⁸ In this example the challenge posed by fine individuation can be assumed away.

²²⁹ It is worth noting that to some this statement may not be uncontroversial. For example, in *Weighing Lives*, Broome claims that the 'better than' relation, the comparative relation for value, is transitive. He does not argue for the claim, he merely asserts that comparative relations, like the betterness relation, are transitive. E.g., "That comparative relations are transitive is self-evident. It is an axiom that lies at the foundation of our arguments. It does not itself need to be supported by argument, and not much argument is available to support it directly." (pg 52.)

As is the case with many unsupported assertions, the force of Broome's claim is questionable. If the earlier example is not persuasive, consider this simple example. Assume that there is a plurality of value. Specifically, for present purposes assume value has three dimensions designated A, B and C respectively. These values are manifest in three candidates X, Y and Z to different extents. This can be represented as follows:

	A	B	C
X	10	-	5
Y	5	10	-
Z	-	5	10

Here the higher number the more valuable the candidate has in that dimension, and the dash represents that the candidate does not exhibit value of this kind.

In this situation, X is more valuable than Y, Y is more valuable than Z, and Z is more valuable than X. Or X is better than Y, Y is better than Z, and Z is more valuable than X. The betterness relation is intransitive.

Proponents of Broome's view may argue that this is not one comparative relation, but three different ones. And to a certain extent that is a natural enough way to read things. But this is exactly point of acknowledging that there may be a plurality of values. If this is the case, then the betterness relation may reflect betterness along multiple dimensions. And as a consequence, the betterness relation may be intransitive.

reasons may recommend having intransitive preferences. If Michael rationally formed his preferences based on the reasons that there are, Michael's preferences would be intransitive.

A successful account of preferences, it would seem, should offer the basis for justifying the transitivity condition in situations in which it applies, such as situations in which a single type of value is relevant. And it should not offer the basis for justifying the transitivity condition in situation in which it does not apply, such as situations in which value itself is intransitive because there is a plurality of value. This nuance can be incorporated into the goal for this section. For example: Offer a justification for the transitivity condition only in situations in which the transitivity conditions is a condition of rationality.

Since it has already been shown that the suggested view can justify the transitivity condition in situations in which it does apply, it only remains to show that it does not justify the transitivity condition in situations in which it does not apply. To see this, merely consider a case in which value is intransitive, and the balance of reasons recommends having intransitive preferences. The agent has true beliefs about the elements of the reasons that there are, recognizes the reasons as reasons and gives each of the reasons an appropriate role in deliberation. If the agent deliberated correctly, then the agent's preferences

Broome may seek to argue that there are not a plurality of values. Or that all values are commensurable. If either of those are the case, then value would be transitive. Two points are worth mentioning about this. First, the question of whether or not there are a plurality of values is quite different then the question of whether it is self-evident that comparative relations are transitive. There would be reasons for concern if one relied on conclusions about the former based on assertions regarding the latter.

Second, if it turns out that there is not a plurality of value, all the better for the suggested view. It can offer the justification of the transitivity condition discussed in the previous section. The point of the current discussion was to recognize the real possibility that there are a plurality of values, to point to the complexities this would create for the suggested view and acknowledge its limitations.

would be intransitive; and rationally so, for he satisfied the fundamental requirement of rationality.²³⁰

As a result, the justification of the transitivity condition offered by the suggested view meets this augmented variation of the goal for this section. The suggested account offers a justification of the transitivity condition only in situations where the transitivity condition is a condition of rationality.

While there is more to explore about the suggested view of preferences, having looked at whether the suggested account can offer justifications for the completeness and transitivity conditions as well as meet the challenge posed by the prospect of fine individuation, the current discussion can come to a conclusion.

²³⁰ Note: the mere fact that there is a plurality of values does not mean that an agent's intransitive preferences are rational. The agent is still subject to the fundamental requirement of rationality. It only means that there is a limitation to the applicability of the transitivity condition.

Chapter 5: Conclusion

The fact that there are multiple interpretations of 'preference' creates a challenge for Preference Theories in part because it threatens to undermine the justification for the consistency conditions on which it relies.

Building on the conception of practical reasoning enumerated in previous chapters, this discussion suggested an account of preferences as all things considered judgments. Such judgments have a rationally governed connection to choices and can reflect an agent's assessments of his own welfare and other values that his choices may realize. As such, it was suggested, this account shared many of appealing characteristics of standard interpretations of 'preference', yet avoided the pitfalls of reducing 'preference' to choices or welfare judgments.

Employing this account, the discussion turned to investigate possible justifications for the completeness condition. It noted that the standard argument for justifying the completeness condition, the 'forcing procedure', did not justify the completeness condition as a condition of rationality; and acknowledged that the suggested view of rationality would not fare any better. For an agent may rationally fail to have complete preferences by simply being unaware of the available alternatives.

This prompted the discussion to reconsider the role of the completeness condition in Preference Theories and suggested it bears the hallmarks of a domain assumption. As a domain assumption the justification of the

completeness condition rests on its ability to demarcate those domains in which the theory may apply. The role of the associated conception of practical reasoning in justifying the completeness condition as a domain assumption is to shed light on the characteristics of circumstances in which it will apply. For example, the suggested conception of practical reasoning suggests completeness conditions is a justified domain assumption that would likely obtain in situations in which agents were fully aware of available alternatives and in which a single type of value was at stake.

The discussion next turned to the challenge posed by the prospect of fine individuation. Building on Broome's suggestion that the challenge could be addressed by stipulating a Principle of Justification by Justifiers, PIJ, the discussion introduced the Normative Identity Condition to justify a variation of Broome's articulation, namely PIJ'. PIJ' stated that alternatives should be individuated from one another, if and only if they differ in terms of relevant recognized reasons.

While PIJ' addresses the challenges posed by the prospect of fine individuation, it also brought to light a noteworthy characteristic of the current account. On the current account preferences are intensional. Preferences are judgments of the agents, and as such it is the agent's understanding of their judgments, the alternatives they are choosing over, and the reasons for doing so which matter. As a result, in some instances convicting an agent of an irrationality, such as having intransitive preferences, or even understanding what transitive preferences call for, will require understanding the choice situation as the agent sees it. Rather than a drawback of the current account, it was suggested this is

a virtue. For the suggested account offers a principled means for addressing the challenge of fine individuation while acknowledging that what rationality calls for depends in part on the agent's understanding of the choice situation.

With this, the discussion turned to the question of justifying the transitivity condition. It began by stipulating an idealized situation to make the question tractable. It was assumed that only consequentialist first order reasons were relevant and only one type of value was at issue.

Appealing to the conception of practical reasoning developed to this point, the discussion then showed that in these ideal situations cases in which an agent's preferences are intransitive are cases in which the agent fails to give a reason he recognizes an appropriate role in deliberation. This is a violation of the fundamental requirement of rationality. And as a result, in these ideal situations, the transitivity condition is a justified condition of rationality. However, when the idealizing assumptions are relaxed the situation changes.

The discussion quickly noted that in situations in which there is a plurality of value, value itself may be intransitive. And, as a result, in situations in which there is a plurality of value an agent can rationally have intransitive preferences. Further, it was shown that in such circumstances the suggested account did not justify considering intransitive preferences irrational. So the suggested view offered a justification of the transitivity condition only in situations in which it applies.

Cumulatively this chapter and those that came before it suggests an account of preference and a conception of practical reasoning that can serve to anchor Preference Theories by offering a justification of the completeness and transitivity conditions and addressing the challenge posed by the prospect of fine individuation. Yet the picture of Preference Theories that emerges is both broader and narrower than commonly conceived. On the one hand the justifications of the consistency condition offered apply within specified domains. On the other, Preference Theories can be seen as crucial component in a broader conception of how people navigate a normatively complex world and make choices. This has the potential to expand the boundaries to which Preference Theories can fruitfully apply without threatening to reduce one's understanding of new domains to fit the dictates of the formalism.

Bibliography

Anderson, Elisabeth (1993), *Value in Ethics and Economics*, Cambridge, Mass.: Harvard UP.

Bratman, M., *Intention, plans, and practical reason*, Cambridge, Mass. : Harvard University Press, 1987.

Brentano, Franz (1969 [1889]), *The Origin of Our Knowledge of Right and Wrong*, London and Henley: Routledge & Kegan Paul.

Broome, J., "Reasoning", unpublished manuscript, 2005.

Broome, J., "Reasons", in *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*, edited by R. Jay Wallace, Michael Smith, Samuel Scheffler, and Philip Pettit, Oxford University Press, 2004, pp. 28–55.

Broome, J., "Normative practical reasoning", *Proceedings of the Aristotelian Society*, Supplementary Volume 75 (2001), pp. 175–93.

Broome, J., *Ethics Out of Economics*, Cambridge ; New York : Cambridge University Press, 1999.

Broome, J., "Normative requirements", *Ratio*, 12 (1999), pp. 398–419. Reprinted in *Normativity*, edited by Jonathan Dancy, Blackwell, 2000, pg. 78–99.

Broome, J., "Can a Humean be moderate?", in: Frey, R G (ed), *Value, Welfare, and Morality*, Cambridge University Press : New York, 1993.

Broome, J., *Weighing Goods: Equality, Uncertainty and Time*, Oxford : Cambridge, Mass. : Basil Blackwell, 1991.

Chisholm, Roderick M. (1986), *Brentano and Intrinsic Value*, Cambridge: Cambridge UP.

Crisp, R., 'Value, Reasons, and the Structure of Justification: How to Avoid Passing the Buck' in *Analysis* 65, 2005, pp. 80-85.

Davidson, D. "Paradoxes of Irrationality", *Problems of Rationality*, Oxford : Clarendon Press ; New York : Oxford University Press, 2004.

Davidson, "Actions, Reasons and Causes", *The Journal of Philosophy*, Vol. 60, No. 23, American Philosophical Association, Eastern Division, Sixtieth Annual Meeting (Nov. 7, 1963), pp. 685-700.

Davidson, D., McKinsey, J., and Suppes, P. (1955) "Outlines of a Formal Theory of Value, I" *Philosophy of Science* 22, pp. 140-160.

Debreu, G. *Theory of value : an axiomatic analysis of economic equilibrium*, New Haven : Yale University Press, 1959.

- Ewing, A. C. (1947), *The Definition of Good*, London: MacMillan.
- Fehige, C. and Wessels, U., *Preferences*, Berlin ; New York : W. de Gruyter, 1998.
- Gibbard, Allan (1998), "Preference and Preferability" in Christoph Fehige and Ulla Wessels (eds.), *Preferences - Perspectives in Analytical Philosophy*, Berlin: Walter de Gruyter.
- Hausman, D., *Preference, Value, Choice, and Welfare*, Cambridge University Press, NY, NY, 2012.
- Hausman, D. "Sympathy, Commitment and Preferences", *Economics and Philosophy*, 2005, 21, pp. 33-50.
- Hicks, J. R., "A Revision of Demand Theory", Oxford, Clarendon Press, 1956.
- Heuer, U. "Beyond Wrong Reasons: The Buck-Passing Account of Value" work in progress.
- Heuer, U. "Explaining Reasons: Where Does the Buck Stop?", *Journal of Ethics and Social Philosophy* 1 (3):1-25, 2006.
- Kolodny, N. "Why Be Rational", *Mind*, New Series, Vol. 114, No. 455 Jul., 2005, pp. 509-563.
- Levi, I, *Hard Choices: Decision making under unresolved conflict*, Cambridge ; New York : Cambridge University Press, 1986.
- Mandler, M. "A Difficult Choice in Preference Theory: Rationality Implies Completeness or Transitivity but Not Both" in *Varieties of practical reasoning* / edited by Elijah Millgram, Cambridge, Mass. : MIT Press, 2001.
- McDowell, John (1985), "Values and Secondary Qualities" In Ted Honderich, ed., *Morality and Objectivity*, London & Boston: Routledge & Kegan Paul, pp. 110-129.
- Musgrave, A., "'Unreal Assumptions' in Economic Theory: The F-Twist Untwisted", *Kyklos* 34, pp. 377-387 August 1981.
- Parfit, *On What Matters*, Oxford University Press, 2011.
- Piller, C., "Normative practical reasoning", *Proceedings of the Aristotelian Society*, supplementary volume 75, 2001, pp. 195–216.
- Rabinowicz, W and Rønnow-Rasmussen, T, "The Strike of the Demon: On Fitting Pro-Attitudes and Value" *Ethics* 114 (3), 2004.
- Rawls, J., *A Theory of Justice*, Cambridge, Mass.: Harvard UP 1971.
- Raz., J. "Reasons: Explanatory and Normative", In Constantine Sandis (ed.), *New Essays on the Explanation of Action*. Palgrave Macmillan, 2009.

Raz, J. "The Myth of Instrumental Rationality", *Journal of Ethics & Social Philosophy*, Vol. 1, No. 1, April 2005.

Raz, J. Practical Reasons and Norms, London : Hutchinson, 1975.

Samuelson , P. A. "The Problem of Integrability in Utility Theory", *Economica*, v. 17, 1950, pp. 355-385.

Samuelson, P. A., "Consumption Theory in Terms of Revealed Preference", *Economica*, v. 15, 1948, pp. 243-253.

Samuelson, P. A., "A Note on the Pure Theory of Consumer's Behaviour", *Economica*, v. 5, 1938, pp. 61-71.

Scanlon, "Structural Irrationality" in Common Minds: themes from the philosophy of Philip Pettit, Oxford : Clarendon Press ; New York : Oxford University Press, 2007, pg 84–103.

Scanlon, T. M., What We Owe to Each Other, Cambridge: Harvard University Press 1998.

Schroeder, M., 'Buck-Passers' Negative Thesis' in *Philosophical Explorations* 12(3), 2009, 341-47.

Sen, A.K., Choice Welfare and Measurement, First Harvard University Press, 1998.

Sen, A.K., "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," in *Philosophy and Public Affairs*. Vol. 6, no. 4, 1977, pages 317-344.

Sen, A.K., On Economic Inequality, New York, Norton, 1973.

Väyrynen, P., 'Resisting the Buck-Passing Account of Value', in Russ Shafer-Landau (ed), *Oxford Studies in Metaethics*, vol 1, 2006, 295-324.

Wiggins, D., "A Sensible Subjectivism?", essay V in D. Wiggins, *Needs, Values, Truth: Essays in the philosophy of value*, Oxford: Blackwell, 1987.

Williams, B. "Internal and External Reasons", *Moral Luck*. Cambridge: Cambridge University Press, 1981.