October 2015

Rethinking Empathy: Value and Context in Motivation and Adaptation

O'Neal Buchanan
The University of Western Ontario

Supervisor
Gillian Barker
The University of Western Ontario

Graduate Program in Philosophy

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

RETHINKING EMPATHY: VALUE AND CONTEXT IN MOTIVATION AND ADAPTATION

(Thesis format: Integrated Article)

by

O'Neal Buchanan

Graduate Program in Philosophy

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

The broad aim of this dissertation is to present an alternative approach to empathy research. The three main questions raised are: What is empathy? How do its component psychological processes become active and operate together? How did empathy evolve? In answering these questions, most researchers have started from a conventional approach that can be described as focusing on short-term phenomena "inside the head" of an individual, evidence that is gathered exclusively in a laboratory environment, and neurocognitive processes that are universally shared by all humans.

A problem with the conventional approach is that it makes social and normative issues in empathy research very difficult to analyze, let alone resolve. Issue such as: which contextual and social variables affect whether empathy occurs or not? Why do conflicting assessments of when empathy has occurred arise? And how should we decide between them? Resolving these issues on the conventional approach is hard because it ignores many important variables affecting empathy. There is a need for rethinking empathy in way that will more fully integrate the normative and environmental variables that affect its interactive complexity.

My approach integrates three such variables. Namely, the environmental contexts of agents and targets, their values, and the motivational stances they adopt towards each other. In the first paper, I attempt to sort out disagreements about what empathy is. I then argue for an account of empathy that emphasizes care and other values. In the second paper, I explore the consequences of unconscious emotions for an account of empathy's theory of empathic accuracy. In the third paper, I criticise evolutionary psychological accounts of empathy, and propose an enlarged theoretical framework that renders the account I develop in the first two papers consistent with recent advances in evolutionary theory, computer science, evolutionary biology, and philosophy.

The three papers of this dissertation present a more detailed causal explanation of empathy that enables us to better predict when it will occur. It sketches a new evolutionary explanation of empathy. And it creates a theoretical space for the analysis of normative and environmental variables that will arguably be important for studying empathy going forward.

Keywords

Empathy, Empathic Accuracy, Care, Values, Motivation, Unconscious Emotion, Evolution, Adaptation, Evolutionary Psychology

Acknowledgments

I wish especially to thank my dissertation advisor, Professor Gillian Barker, for challenging me through her teaching, for being an inspirational academic, and a kind mentor; my parents for their love and support; Professor John Nicholas for his thought-provoking and joyful classes; Professor Fred Adams for his generosity and philosophical guidance; Professor Eric Desjardins and Professor Peter Williamson for their insightful and helpful comments in the final stages of this project; and Professor Louis Charland and Professor Angela Mendelovici for their comments in its early stages. I would also like to thank the support staff and the professors of Western University's Philosophy Department and Rotman Institute of Philosophy. And also, the many friends and colleagues with whom I have discussed this project and philosophy over the years. Among them are: Sean Coughlin, Riin Sirkel, Ryan Middleton, Rodney Parker, David Burty, and Alexandre Sannen.

This dissertation is dedicated to my mother, Alice Buchanan (1948-2014).

Table of Contents

Abstract ii

Table of Contents v

List of Tables ix

List of Figures x

List of Appendices xi

1. Introduction 1

    1.1 How does empathy work, and when does it occur? 1

    1.2 Why is empathy important and philosophically significant? 2

    1.3 What is my approach to empathy and how is it different from the conventional approach? 6

    1.4 What is the need for an alternative approach to empathy? 9

    1.5 Paper 1: The care account and the role of values in empathy 10

    1.6 Paper 2: Two challenges to empathy from unconscious emotions 11

    1.7 Paper 3: Towards and enlarged evolutionary psychological explanation of empathy 13

    1.8 What are the future directions in empathy research? 15

2. The care account and the role of values in empathy 24

    2.1 Introduction 24

    2.2 The phenomenon of empathy 26

    2.3 Various uses of ‘empathy’ 28

    2.4 The relationship between what empathy is and empathic accuracy 33

    2.5 Two types of empathy accounts 36

        2.5.1 Matching accounts 37

        2.5.2 Valence 39

        2.5.3 Concern accounts 40

2.6 Empathy as a motive 42
2.7 Care in empathy 45
2.8 The causal roles of care 46
2.8.1 Care: before *involuntary* perspective-taking 47
2.8.2 Care: after *involuntary* perspective-taking 48
2.8.3 Care: before *voluntary* perspective-taking 49
2.8.4 Care: after *voluntary* perspective-taking 50
2.9 On the separation of care and empathy 51
2.10 On the constitutivity of care in empathy 59
2.10.1 Mindreading 60
2.10.2 Emotional contagion 62
2.10.3 Mimicry (behavioural and neural) 63
2.10.4 Like-me perspective-taking 64
2.10.5 Like-other perspective-taking 66
2.10.6 Summarizing constitutivity 68
2.11 The roles of "values" in empathy 70
2.12 Care and other values 70
2.13 A care account and empathic accuracy revisited 73
2.13.1 Values influence what causes empathy (i.e. they are triggering causes) 74
2.13.2 Values influence why empathy occurs (i.e. they are structuring causes) 79
2.13.3 Values influence how empathy operates (i.e. which motives are involved) 81
2.14 Concluding remarks 85
3. Two challenges to empathy from unconscious emotions 95
3.1 The two challenges 95
3.2 Two types of accounts: *matching* and *concern* 96

3.3 Unconscious emotions 99
3.4 Empathy in three parts 101
3.5 Challenge I: Unconscious matching 105
3.5.1 First reply: Eliminativism 107
3.5.2 Second reply: Restrictivism 109
3.5.3 Third reply: Rejecting the challenge 112
3.6 Challenge II: Unconscious concerns 116
3.6.1 The indeterminacy of affect valence reply 119
3.7 Concluding remarks 123
4. Towards an enlarged evolutionary psychological explanation of empathy 130
4.1 An enlarged evolutionary psychology of empathy 130
4.2 Standard evolutionary psychology [SEP] of empathy 133
4.3 The evolutionary framework of SEP: Three tenets 142
4.3.1 Inclusive fitness 143
4.3.2 The mind is like a computer 145
4.3.3 Evolution is slow 147
4.4 Mutually supportive tenets 149
4.5 The method of SEP 150
4.6 An enlargement of theory 154
4.6.1 Niche construction and mutualism 157
4.6.2 The mind is like a self-adaptive computer 162
4.6.3 Evolution is faster than we thought: feminist evolutionists and biological leverage 166
4.7 A mutually supportive theoretical enlargement 170
4.8 The revised method of EEP 173
4.9 Enlarged evolutionary psychology of empathy 183

4.10 Concluding remarks 187

Historical Appendix 196

Curriculum Vitae 199

List of Tables

Table 1: Component processes of empathy compared to typicality and necessity of care..... 68

Table 2: The method of SEP and EEP 181

List of Figures

Figure 1: Motivational orientations and terminal values 78

Figure 2: Motivational orientations and instrumental values 83

Figure 3: An agent matching a target's experienced emotions outside of awareness 112

Figure 4: Mutually supportive tenets of SEP 149

Figure 5: Mutually supportive tenets of EEP 173

List of Appendices

Historical Appendix 196

1. Introduction

1.1 How does empathy work, and when does it occur?

The three papers of this dissertation present an *alternative approach* to analyzing and explaining empathy. Imagine that you are walking down the street and someone standing by a shop asks you for money. As you look into their eyes you feel their sadness, their pain, their disappointment. You are aware of their concern for money. And you feel motivated to help them. What is empathy in such a case? How do the psychological processes that bring it about become active and work together? And how did empathy evolve? The alternative approach that I develop seeks to answer these questions in a new way.

The majority of accounts that answer these question start from a *conventional approach* that is characterized by the following features. The conventional approach focuses on short-term phenomena "inside the head" of an individual agent. It appeals to evidence that is gathered exclusively in a laboratory environment. And it posits neurocognitive mechanisms or processes that are (universally) shared by all humans. The conventional approach to empathy needs to be rethought because it does not sufficiently take into consideration important aspects of empathy. Namely, the v*alues* of both agents and targets that influence how and when it occurs; their *motivational orientations*; and how empathy is affected by changes in *environmental contexts*.

Conventionally, empathy researchers pay little attention to these factors. They mostly answer the question of how empathy occurs by focusing on the nature or format of the mental processes of the empathy agent (Goldman 1992; de Vignemont and Singer 2008). I argue that researchers taking the *conventional approach* do not explain empathy in sufficient detail, and that they are unable to predict when it is more or less likely to occur. One reason for this is that, depending on the model, there are over a dozen different phenomena to which the term "empathy" refers. And though some researchers have begun to discuss the resulting conceptual difficulty (Batson 2009), most seem to regard

this divergent usage as merely terminological, and therefore unimportant. But indeed this variation is of great importance because different uses of the term "empathy" have different consequences for what counts as *empathic accuracy*. Models of how empathy works all employ the notion of *empathic accuracy*—which specifies the end-state criteria for an empathizer to "get it right". Accordingly, the question of *how* empathy works is premature because the question of *what* empathy is has not been satisfactorily answered (as is apparent in the diverse ways that *empathic accuracy* is used).

On my model, empathy is more likely to occur when agents are in a *cooperative* or *altruistic motivational orientation* towards targets, and when targets are in a cooperative or altruistic motivational orientation towards agents. A motivational orientation is an agent's disposition to "define, compare, and evaluate their behavioural alternatives not only in terms of their implications for achieving [their] own preferred ends, but also in terms of their implication for the outcomes that will be afforded to others." (McClintock 1972, p 438). I argue that both agents and targets are more likely to be in cooperative or altruistic orientations when they *care for* or *value* each other's well-being. To more fully explain how empathy works and to predict when it will occur, we should consider motivational orientations that are associated with values, and how these orientations are affected by changes in environmental contexts.

1.2 Why is empathy important and philosophically significant?

We often think of empathy as sharing people's emotions, taking their perspective, and feeling motivated to help them with their concerns. To take another example, say a friend tells me that they were recently fired from a job that they found fulfilling. I may share their disappointment, their belief that they were dismissed unjustly, and I may feel motivated to help them find a new job. I empathize with my friend by experiencing emotions and beliefs similar to theirs, and by being motivated to help them in a manner appropriate to their concerns. Accounts of empathy usually include three component mental processes: *emotional contagion*, *mimicry*, and *perspective-taking* (Preston and de

Waal 2002; Stotland 1969).[1] These processes contribute to an empathizing agent having similar psychological states as those of a target, and to an agent becoming aware of a target's concerns. Some accounts also take empathy to include an additional *motivational process* that can cause an agent to help a target (Sagi and Hoffman 1976; Batson and Shaw 1991).

There exist many philosophical and scientific accounts of empathy (Hoffman 1987; Darwall 1998; Batson 1991; Gallese 2001; Prinz 2011a; Zahavi 2011). This interest is unsurprising because empathy has been shown to be important to many social practices and capacities. It is taken to be a key contributor to fundamental human social capacities such as moral judgment and conflict resolution, and it also deemed essential to many more particular social practices such as interviewing, field work, journalistic reporting, clinical work, social work, legal judgment, advertising, and so on. It plays an explanatory role in many scientific theories such as in psychological and biological theories of prosocial motivation (Krebs 1970, 1975; Hoffman 1981; de Waal 1996; Sober and Wilson 1998; Wilson 1998; Richerson and Boyd 2005; Newson and Richerson 2013).[2] And in environmental ethics, it has been argued that empathy makes it possible to understand that nonhuman organisms have intrinsic value and should, for this reason, be preserved (Callicott 1986). Examining these practices and theories reveals that what empathy is taken to be varies. For example, in psychiatry and psychology, a lack of empathy is said to explain both autism and psychopathy (Kennett 2002). But clearly this lack of empathy results in two very different classifications. This is partly because researchers do not use the notion empathy consistently across all theories of autism and psychopathy. It is also because they have not made explicit many important factors that influence how empathy works differently in each case.

Empathy also figures prominently in many historical and contemporary philosophical debates. The word 'empathy' was translated into English by the psychologists Edward B. Titchener in 1909 from the German word '*einfühlung*'. The latter was coined by Robert Vischer (1873) in his doctoral dissertation in philosophy. In the context of that period's

[1] A detailed description of these processes and their roles in causing empathy is in paper 1.
[2] Although, in the cases of Sober and Wilson (1998), the explanatory focus is on altruism.

philosophy and psychology, ‘einfühlung’ and ‘empathy’ were technical terms in debates about the psychology of aesthetic experience that referred to the processes by which an agent felt their way into a work of art in order to discover and create conscious states within it.[3] Eventually, it became used to refer to a process occurring primarily among humans (as opposed to among humans and art objects or humans and non-human animals). It was around this time that discussions of empathy became important in debates about the *problem of other minds*—the problem of whether or not and how we know that others have minds and how we know their mental states (Coplan and Goldie 2011, xiii). This is important because the side on which one fell in these debates had serious implications for how to study the mind. Namely, if a researcher believed that it was not possible know the mental states of targets, then that researcher might, by entailment, be committed to *behaviourism*. On the other hand, if a research believed that it was possible, by empathizing, to become aware of the mental states of targets, then that research might be committed to *mentalism*.

More recently, in the 1980’s and 90’s, empathy was again prominent in debates about how we attribute mental states to others (Baron-Cohen et al. 1985; Stich and Nichols 1992; Goldman 1992; Gopnik and Wellman 1995; Baron-Cohen 1995). The issue here was about the cognitive format of the mental processes allowing us to make such attributions. Researchers argued that we either *simulated* the mental states of a target in a mental format akin to sensory imagination, or we created a *theory* of their mental states in an amodal format. In these debates, empathy was equated with *mindreading*—“the capacity to understand other minds” (Nichols 2004, p 8). Accordingly, it was often mistakenly treated as being a process that does not involve any shared emotional experience between agents and targets. Once this had been corrected, the earlier debates about mindreading became important to debates about the metaphysics of empathy. *Simulation theorists* and *theory theorists* agree that our attribution of mental states to others *and* our ability to empathize with them depend entirely on processes that are *representational*—that is, on processes whereby an agent attempting to empathize with a target generates a representation of that target’s mental states. Empathy occurs then in

[3] For more details, see Appendix A.

virtue of the agent accessing representations (either simulations or theories) internal to their own mind. Contrary to this, other theorists argue that, when empathizing, agents have *direct-perceptual access* to the minds of targets (Zahavi 2001, 2011; Gallagher 2008, IP). These latter theorists believe that we not only have direct access to our own minds, but that we also have direct access to the minds of others in our perceptions them. I describe these competing views in more detail in the first paper. I refer to them here simply to show some of the significance of empathy for philosophical debates—in this case, for debates about the metaphysics of mind, mindreading, and empathy.

As mentioned, the term 'empathy' originated in philosophical debates about aesthetics. And empathy featured prominently in debates about the problem of other minds; and more recently, in debates about mindreading and the metaphysics of mind. Also, how theorists think about empathy has significantly affected at least two other topics of philosophical concern: ethics and the evolution of altruism. In ethics, Slote has developed an "ethics of empathic caring" according to which "empathy is the primary mechanism of caring, benevolence, [and] compassion" (Slote 2007, p 4). He argues that empathy and cultivating our ability to empathize is central to our moral education and development (*Ibid.*). In the feminist tradition, Noddings argues that empathy contributes to self and other-understanding, which in turn fosters ethically healthy appreciation and criticism (Noddings 2002a, p 153; 2002b; 2003).

With regard to the evolution of altruism, Sober and Wilson claim that empathy evolved to cause agents to behave altruistically (Sober and Wilson 1998). They favour the theory of *psychological altruism*—the view that organisms have evolved to sometimes act with an *ultimate concern* for others—as opposed to the theory of *reciprocal altruism*—the view that organisms have only evolved to act with a concern for others in case of a likely returned benefit (Sober and Wilson 1998, p 7). Their account is a scientifically informed philosophical contribution to the longstanding debate about whether humans, by nature, are essentially *psychological egoists* or *psychological altruists*. In this debate, Sober and Wilson take empathy to be the process that is primarily responsible for causing agents to behave altruistically (Sober and Wilson 1998, Ch. 8). They argue that empathy causes a motivation to help targets with their concerns–and that this motivation is often altruistic

in that its intended end is to benefit the target (*Ibid.*). But the most influential account of the evolution of empathy is in evolutionary psychology.

Evolutionary psychologists claim that there exist essential sex differences in the frequency and accuracy of empathy, and that these differences are caused by innate mechanisms that have remained fixed since the distant evolutionary past (Baron-Cohen 2003; Pinker 2011). In the past, empathy is believed to have functioned to solve adaptive problems such as "mothering", "gossip", "social mobility", and "reading your partner" (Dunbar 1998; Baron-Cohen 2003, p 425-437). And because these functions of empathy are purported to have been mostly performed by females throughout evolutionary history (until perhaps very recently), evolutionary psychologists claim that this resulted in the continued selection for genes causing psychological mechanisms that explain contemporary differences in empathy (Baron-Cohen 2003; Baron-Cohen and Wheelwright 2004; Lawrence et al. 2004).

1.3 What is my approach to empathy and how is it different from the conventional approach?

As mentioned, many sources of empirical evidence have recently shaped empathy research. An account that draws from these sources is the *mirror neuron account* of empathy (Gallese 2001; 2002). It also is a prominent example of an account that follows the conventional approach. Mirror neurons fire both when we perform an action, and when we observe another organism performing a similar action. Their activation is recorded in hand-gesture experiments over short time periods in a laboratory setting. Mirror neurons are taken by many to be the basis of our ability to *simulate* the internal states of others (Gallese 2001; Rizzolatti and Sinigaglia 2008)—and thus, to fully explain how an agent empathizes with a target (Gallese, Keysers, and Rizzolatti 2004).

A major problem with the conventional approach is that it makes social and normative issues in empathy research very difficult to analyze, let alone resolve. Issue such as: which *contextual* and *social variables* affect whether empathy occurs or not? Why do *conflicting assessments* of when empathy has occurred arise, and how should we decide between them? And what is the role of *values*, and *motivations*? Addressing these issues

will be important for empathy to play a role in resolving conflicts and for it to continue to be seen as a source of moral good. To begin resolving these issues we should consider how agents and targets of empathy interact in varying environments. But the conventional approach avoids them. It is too narrowly focused on an agent's neurocognitive mechanisms during short time periods in a laboratory environment, and on the purported species-wide instantiation of such mechanisms. For example, an account of empathy following the conventional approach will have difficulty analysing the empathic experience of what Hoffman calls "complex combinations":

> A shabbily dressed man is observed robbing an obviously affluent person on the street. A young child might feel empathic and sympathetic distress for the victim and anger at the immediate, visible culprit. Mature observers might have these same feelings, but a variety of other empathic affects as well. They might feel guilty over not helping the victim. If they are ideologically liberal, they might empathize and sympathize not only with the victim but also with the culprit because of his poverty. The observers might view the culprit as a victim of society and feel empathic anger towards society. Furthermore, if the observers are affluent as well as liberal, they might feel guilty over being relatively advantaged persons who benefit from the same society. Ideologically conservative observers might not sympathize with the culprit but might respond with unalloyed empathic anger instead. They might also feel empathic anger toward society, but in this case because they view the victim, not the culprit, as a victim of society (because of inadequate law enforcement and citizen protection) (Hoffman 1984, p 657).

Indeed all occurrences of empathy are *complex combinations* because all occurrences of empathy are importantly influenced by both the values of agents and targets, as Hoffman's example makes clear. But accounts following the conventional approach do not show how values influence the way empathy works and when it occurs. They do not, for example, consider the relationships between emotional states and motivational orientations that are associated with differences in values.

The alternative approach to empathy that I develop in this dissertation integrates three variables: *values*, *motivational orientations*, and *environmental contexts*. And the dynamic interactions between these variables and other contextual factors is brought to the fore. I consider the role of *terminal values*—qualities of outcome states that are desirable; and the role of *instrumental values*—qualities of contextually appropriate

modes of conduct (Lovejoy 1950). And I include four motivational orientations: *indifference*; *competition*; *cooperation*; and *altruism* (McClintock 1972). These motivational orientations are steered by the goals that agents and targets bring to a situation. And each motivational orientation is associated with different values which in turn affect their occurrent beliefs and emotions, depending on their environmental context (including other organisms). The relationship between goals, motivational orientations, values, and contexts can be used to assess whether or not empathy has occurred, and to predict whether or not it will occur. I argue that this is because when values change, motivational orientations towards targets may also change and vice versa. Values also affect the content of the concerns of both agents and targets of empathy. Thus, it is important to understand the role of values in an agent's ability and likelihood to empathize with a target. And it is important to understand the role of values in how targets of empathy understand their own situations. A more detailed account of empathy that includes the dynamic interactions between goals, motivational orientations, values, emotions, and environments emerges from the alternative approach that I develop in these three papers.

Empathy is a fundamentally social phenomenon. But most accounts take the conventional approach to explaining empathy by focusing on agents in laboratory environments, often in isolation from other organisms, and even in isolation from elements of their mental lives, such as their ethical beliefs and commitments, that importantly affect their interpersonal behaviours. There is a need for an alternative approach to empathy that is scientifically informed, but also more socially and thus more philosophically informed. Such an approach should be good for understanding the relational nature of empathy; debates in empathy research about our ability to take the perspective of many organisms; and the implications and consequences that these debates have for current socially and ethically important challenges.

1.4 What is the need for an alternative approach to empathy?

Widespread interest in empathy has increased in the last 15 years. Some of this interest is due to economic and technological change. In this time, trade and communication have become even more technologically mediated. These transactions are now faster and more global. Paradoxically, in the midst of this communication revolution, group tensions have taken center stage in policy deliberations and in the media. Preventing and combatting international terrorism is now at the top of the political agenda in Canada and many other countries. And while I write this introduction, people motivated by a desire for equality and justice across racial identities are protesting in 170 U.S. cities. A fuller account of and use of empathy may contribute to progress on these and related issues (Calloway-Thomas 2010).

Another source of the renewed interest in empathy is the emerging view of human nature in biology and studies of the mind. Hobbes's view of human nature is well known: humans are essentially self-interested and the world is essentially competitive. This view of human nature and evolution has been popularized by Dawkins who states that: "'nature red in tooth and claw' sums up our modern understanding of natural selection admirably" (Dawkins 1989). More recently however, researchers have begun to develop theories according to which humans are by nature more cooperative and altruistic than previously thought (Sober and Wilson 1998; de Waal 2010; Rifkin 2010). Along these lines, the discovery of mirror-neurons that contribute to our interpretations of others, the discovery of empathy's role in autism and psychopathic disorders, as well as developments in cognitive science, ethics, and the philosophy of mind and psychology are contributing to a renewed sense of optimism about empathy.

But these advances in research (and the resulting optimism) are not without their theoretical difficulties and detractors. Confusion about what empathy is has become a significant impediment to further advances in empathy research. As Batson (2009) and Pinker (2011, Ch. 9) point out, the word 'empathy' refers to many different mental states

and phenomena.[4] And some theorists such as Prinz (2011a, 2011b) have argued that empathy is not beneficial to ethical theory or to the resolution of large-scale ethical challenges such as environmental destruction or disease relief. An alternative theoretical approach to empathy should be worthwhile for empathy research to continue benefiting the many domains it does, and for the phenomenon of empathy to contribute to meeting the difficult social challenges we face.

1.5 Paper 1: The care account and the role of values in empathy

The focus of my alternative approach motivates the main questions in each of the three papers. The first paper asks: what is the role of values in empathy? In answer to this question, I present of a novel account of empathy called the *care account*. Care is a value. We care about what happens, about how it happens, about objects, and organisms. In empathy, the value of care is important. Care causes empathizing, and empathizing causes care. When an agent cares for or values the well-being of a target, that agent is more likely to empathize with that target. This is because the value of care is most often subsumed under the motivational orientations of cooperation and altruism. When an agent is engaged in an activity that involves or requires cooperation or altruism, that agent's behaviours can be partially explained by a motivation that is informed by the value they place on the target's well-being. This in turn provides some insight into the types of environmental contexts that an agent is more likely to empathize in. For example, in a competitive environment such as a meeting with business rivals, empathy is less likely than at a dinner table with friends. Care also results from empathy. An agent does not necessarily have to care for a target prior to attempting to empathize with that target. But I think that when an agent empathizes accurately with a target, that agent cares for the target; this is because (as I suppose) empathy involves an agent becoming aware of a targets' concerns and feeling motivated to help the target with those concerns. Thus,

[4] For example, some researchers take 'empathy' to refer to an agent having knowledge of the emotional state of a target. Others take it to involve an agent having the same experience as that of a target. I will provide more examples of the different uses of the term and distinguish between account types in my first paper.

my care account begins by examining the role of care in empathy, which in turn opens up a space for examining of the role of other values in different environmental contexts.

1.6 Paper 2: Two challenges to empathy from unconscious emotions

The main question of my second paper is: what relationship must there be between an agent and a target's *unconscious emotional states* for empathy to occur? The care account I develop in the first paper is a token of the *concern account type* of empathy account. As such, for empathy to occur, it requires a *matching relation* between an agent and a target whereby an agent experiences an emotion with the same *valence* as that of the target (either positive or negative). For example, empathy may occur even if an agent is experiencing pride and a target is experiencing joy, just as long as both are experiencing emotions that are positively valenced. On the other hand, *matching accounts* of empathy have a stricter matching relation requirement for when empathy occurs.[5] On matching accounts, the matching relation between an agent and a target specifies that both must experience the same emotional content. For example, when a target of empathy is experiencing anger, for empathy to occur, an agent must also experience anger.[6] The matching relation required between an agent and target's *conscious emotional states* for empathy to occur is precisely specified on each account type. However, the matching relation required for empathy to occur between an agent and a target's *unconscious emotional states* has not been examined. Clarifying the matching relation required between the unconscious emotional states of agents and targets on each account type of empathy is the main goal of my second paper.

Including unconscious emotions in an account of empathy has important implications for what counts as accurate empathy that have not hitherto been investigated. I draw out these implications by presenting a separate challenge to each account type of empathy: an *unconscious concerns challenge*, and an *unconscious matching challenge*. Exploring each challenge allows me to more precisely define what *accurate empathy* is on each account

[5] Although matching-accounts do not usually discuss the notion of valence.
[6] And not say, sadness, which is also often a negatively valenced emotion.

type.[7] After doing so, I present new possible replies to each challenge from each account type.

Because my care account of empathy is a concern account, the most important reply that I present is the *indeterminacy of affect valence* reply to the *unconscious concerns* challenge. This challenge asks: must an agent become aware of a target's unconscious concerns for accurate empathy to occur? For example, evolutionary psychologists argue that *conscious emotions* can cause *unconscious concerns*. In particular, Cosmides and Tooby (2000) state that the experience of fear causes changes in behaviour that are motivated by concerns that need not be conscious. When an agent walks alone at night, that agent experiences fear because they are possibly being stalked or about to be ambushed (Cosmides and Tooby 2000, p 3). In this case, the agent consciously experiences fear and is consciously concerned about being stalked or ambushed. But with this conscious experience of fear come unconscious concerns such as: a concern for safety; a concern for the location of loved ones; a concern for the location of others than can protect me; a concern for finding a defensive position (*Ibid.*). The questions then are: for accurate empathy to occur on a concern account, is it sufficient that an agent become aware of the target's *conscious concern* about being stalked? Or must an agent also become aware of the target's *unconscious concerns* such as their concern of locating loved ones and finding a safe location? I present a reply to this challenge that draws from Charland's (2005) *indeterminacy thesis of affect valence*. Charland's thesis argues that unconscious emotions are not functionally and behaviourally identical to conscious emotions. Specifically, unconscious emotions are not motivational; whereas, concerns are motivational. Therefore, I claim that on a concern account such as mine, an agent need not match a target's unconscious emotions because they do not cause unconscious concerns.

[7] Each account type differs with regard to what it takes empathy to be. This is not the same as differing with regard to what accurate empathy is. The former defines what an agent is attempting to do when empathizing. The latter (empathic accuracy) defines a state that occurs when an agent's attempt at empathy is successful.

1.7 Paper 3: Towards and enlarged evolutionary psychological explanation of empathy

The main questions of the third paper are: why, when, and how did empathy evolve? I present an enlarged evolutionary psychological answer to these questions. Specifically, I examine evolutionary psychological research on empathy. Evolutionary psychologists conclude that the answer to why empathy evolved is that it functioned to improve females' capacity for mothering; caused females to gossip more easily and frequently among themselves; caused females to form social alliances with their male partner's associates more easily; and allowed females to anticipate and attend to the needs of their male partners more easily (Pinker 2002, 2011; Baron-Cohen 2003, 2007). They argue that the mental processes of empathy that motivated females to behave in this way were selected for during the Pleistocene, and they are still operative today.[8] This is consistent with the dictum of standard evolutionary psychology that "our modern skulls house a stone age mind." (Cosmides and Tooby, 1997).

These claims about why and when empathy evolved depend on a specific understanding of how it evolved. This understanding or theoretical framework is standard across evolutionary psychology. Its three central tenets are: 1) inclusive fitness; 2) computational theory of mind; 3) slow evolution. These tenets are consistent and mutually supportive. However, my alternative approach to empathy allows us to notice that the theoretical framework of *standard evolutionary psychology* [SEP] is one that is focused on processes internal to the individual; it does not take into consideration values or motivational orientations; and it treats the environment as stable and fixed rather than examining the variety of environments that organisms inhabit, and the ways that organisms have modified their environments throughout evolutionary history. To make my approach consistent with evolutionary theory, I enlarge the framework of SEP to include the following recent advances in evolutionary biology, computer science, and philosophy: 1) niche construction and mutualism; 2) self-adaptive computer software; 3) feminist evolutionary theory and biological leverage.

[8] The Pleistocene refers to the period between approximately 2.6 million years ago to 12 thousand years ago.

This theoretical enlargement has important methodological implications for SEP. It shows that SEP cannot merely rely on the existence of fixed of mental processes that were selected in the distant past to explain current behaviour. Before doing so, it must justify the purported fixity of such processes by more precisely characterizing their causal contribution throughout history (including the more recent evolutionary history of the last 12 thousand years). These implications take the form of a revision to the method of SEP. Specifically I present a *comparative method* according to which many of the steps in the method of SEP remain the same. For example, according to the new method of *enlarged evolutionary psychology* [EEP], it remains important to form hypotheses about the effects that our mental processes causally contributed to in the environments of the distant past. However, the methodological starting point of EEP is different. It begins by generating hypotheses about the current causal contributions of our mental processes (i.e. their current functions). Doing so allows a researcher looking to explain the evolution of empathy to compare the current functions of empathy with its distant historical functions. The benefit of this method is that it allows us to notice whether those functions have remained the same. Furthermore, if functional changes are discovered, then the revised method of EEP also allows a researcher to start examining and tracking those changes. I begin to apply the revised method of EEP in sketching a new explanation of how empathy evolved. And I conclude, contrary to SEP, that relatively recent environmental modifications and resultant social interactions of evolutionary significance (e.g. *biological leverage*) affected the evolution of the psychological processes currently enabling empathy.

EEP can be situated among the many criticisms of SEP insofar as it argues that the functions posited by SEP (including the functions of empathy) are not the only ones that are evolutionarily plausible (Stotz and Griffiths 2002; Greene 2004; Buller 2005; Liesen 2007). But other criticisms of SEP have not discussed how the practice of evolutionary psychology might change as a result of this claim. EEP goes a step further by presenting specific revisions to the standard method of SEP. These revisions have significant consequences for SEP explanations of empathy and other phenomena that it seeks to explain.

1.8 What are the future directions in empathy research?

There is great potential for additional research on empathy and related issues. A debate that is already being addressed by philosophers which requires more attention is about the metaphysics of the mental processes involved in empathy. The question here is whether we always (or ever) have direct access to the mental states of a target (in perhaps the same way that we have access to our own mental states), or whether we always have indirect access to *representations* of a target's mental states? With a more detailed understanding of the how the mental processes involved in empathy work, progress may be made in this philosophical debate. Another debate that is already being addressed by philosophers is about whether empathy is necessary or important for moral action, progress, and theory.[9] This debate has recently been more widely addressed by an interdisciplinary panel in the Boston Review forum entitled *Against Empathy*.[10] Although not all of the researchers on the panel were "against empathy" the question of whether empathy plays or should play an important role in ethics is worthwhile. More specifically, a question that will be important to further consider is: how and under what conditions is empathy a liability or even dangerous? Some researchers believe that empathy is an intrinsically negative process or an intrinsically negative emotion. They claim that this is because empathy is biased towards those who are similar to us. Rather, they suggest that pure (non-emotional) reason is a better guide to right thinking and action. They further claim that empathically motivated behaviour can be manipulated. For example, there is evidence that legal defendants who show emotions receive lighter sentences than those that do not (Tsoudis 2002). Both of these claims—that empathy is biased, and that it can be used for manipulation—are true. But this does not mean that empathy is not important for moral theorizing, moral development, decision-making, and other practices. For example, voluntary empathy may be used to assess and challenge the biases of automatic empathy. However, it does mean that more research is required to better understand the potential dangers of empathy and to help satisfy the concerns of those who are strongly against its role in ethical theories. This dissertation has

[9] Consensus in this debate is probable but not imminent.
[10] http://www.bostonreview.net/forum/paul-bloom-against-empathy

consequences for these and other debates already underway in philosophy and empathy research more broadly. But it may also be used to address two areas that have not yet been examined by philosophers. These areas are more salient after considering the totality of this document. So I will here be brief.

The first area is empathy testing. Empathy tests are used in the selection, training, and assessment of personnel and clients in many fields including management, medicine, pharmacy, nursing, social work, and teaching. The problem here is that current designs of empathy tests (e.g. "empathy quotient" tests designed by evolutionary psychologists), do not differentiate between whether participants are being assessed on their awareness resulting from empathic perspective-taking or on their ability to apply familiar stereotypes (Lindgren and Robinson, 1953). The question of what exactly empathy tests are measuring remains controversial. Empathy tests usually take one of two forms: a brief group interaction session in a plain room or a text-based self-report questionnaire. In the design of empathy tests, two types of experimental prompts (prompts for empathic perspective-taking and prompts for applying cultural norms) should be more clearly distinguished. For example, on self-report questionnaire based empathy tests, respondent answers sometimes coalesce around certain answers that may have been chosen for their propriety rather than having been empathy induced. For example, in answering the question of whether a respondent was friendly to a target, respondents answer "fairly friendly" very frequently. This occurs regardless of whether their other responses support the conclusion that they are "good empathizers" (*Ibid.*). My approach to the analysis of empathy can contribute to developing tests that better distinguish between answers that are caused by a more abstract understanding of what is appropriate to respond, and answers that result from empathic processes such as perspective-taking and feeling motivated to help a target. At least two guidelines for designing empathy tests follow from my approach which integrates goals, emotions and motivations, and social variables. First, empathy tests should be administered over longer time periods. Currently, empathy tests of both types are short in duration. Second, participants should be assessed in varying environmental and social contexts. Empathy is more or less likely to occur as a result of changes in motivational orientations towards targets which change

as a function of social factors such as time constraints. Accordingly, measuring empathy in various environmental and social contexts will be important for distinguishing between the causes of responses.

Empathy contributes to socially and morally desirable behaviour. This is supported by DSM-based diagnoses of anti-social personality disorder and personality tests specific to psychopathy that measure a reduction or absence of empathy (Blair 1995). Further, correlations are often drawn between the results of general personality tests and personality tests specifically designed to measure "empathic ability" or "empathy quotient" (Lawrence et al. 2004). These correlations support the hypothesis that a lack of empathy leads to poor social adjustment. More precisely identifying what is being measured in empathy tests will allow us to develop tests that more accurately measure the intended construct.

The approach developed in this dissertation can be applied to a second topic that has not yet been examined by philosophers; namely *hot-cold empathy gaps* (Van Boven and Lowenstein 2005). Decision making experiments have recently shown that an agent's predictions of a target's decisions are significantly influenced by that agent's own emotional states. This evidence is important because previous theories of decision making (especially under risk or uncertainty) are *consequentialist* and *cognitive* theories. These theories are *consequentialist* in that they predict that people make decisions based on the assessment of the consequences of possible alternatives. And they are *cognitive* in that this assessment is described as an expectation-based calculus. The emotional states and feelings that people have when making decisions are meaningless or epiphenomenal. In short, previous decision-making theories predicted that agents will aim to maximize their *expected utility* (benefit to themselves). Accordingly, predictions about a target's decisions will also be based on predictions about expected utility.

Evidence from research on hot-cold empathy gaps shows that these decision-making theories are incomplete. An agent's emotional state often causes that agent to make false predictions about their own and other people's emotional states and decisions. For example, when an agent is asked to predict whether they would be willing to participate

in potentially embarrassing activities (such as dancing in front of an audience), their prediction varies as a function of the immediacy of the activity (Van Boven et al. 2004). When the activity is to be performed in the distant future, agents overestimate how willing they are to perform it as compared to when the activity is to be performed nearly immediately. That is, when making a hypothetical rather than a real decision to perform, agents were more likely to predict that they would participate. An agent's emotional state also influenced their predictions about other people's potential decisions. In hypothetical versus immediate (potentially) fearful conditions, agents were more likely to predict that others would also participate in the immediately (potentially) fearful activity. In sum, our own emotional states sometimes cause us to make false predictions about other people's decisions because they blind us from parts of a target's perspective.

The relationship between motivational orientations and values developed in my approach to empathy could be integrated into the models used in these experiments. For example, whether an agent is ethically conflicted about performing a certain action could influence their predictions about whether they and other people would perform that action. Similarly, if performing an action is taken to be threatening to an agent's values or to those of others (either present or not, affected by the action or not), the agent's predictions may also vary. Further, the relationship between values and emotional states under hypothetical and real conditions could be modeled and tested.

References

Baron-Cohen, Simon. *Essential Difference: Male and Female Brains and the Truth about Autism*. Basic Books, 2003.

---. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT press, 1995.

---. "The Evolution of Empathizing and Systematizing: Assortative Mating of Two Strong Systematizers and the Cause of Autism." (2007).

---. *The Science of Evil: On Empathy and the Origins of Cruelty*. Basic books, 2011.

Baron-Cohen, Simon, Alan M. Leslie, and Uta Frith. "Does the Autistic Child Have a 'theory of Mind'?." *Cognition* 21.1 (1985): 37–46. Print.

Baron-Cohen, Simon, and Sally Wheelwright. "The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences." *Journal of autism and developmental disorders* 34.2 (2004): 163–175. Print.

Batson, C. Daniel. "These Things Called Empathy: Eight Related but Distinct Phenomena." *The Social Neuroscience of Empathy*. Ed. Jean Decety and William Ickes. Cambridge, MA, US: MIT Press, 2009.

Batson, C. Daniel, and Laura L. Shaw. "Evidence for Altruism: Toward a Pluralism of Prosocial Motives." *Psychological Inquiry* 2.2 (1991): 107–122. Print.

Blair, R. James R. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." *Cognition* 57.1 (1995): 1–29. Print.

Buller, David J. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. MIT Press, 2005.

Callicot, J. Baird. "On the Intrinsic Value of Nonhuman Species." *The Preservation of Species*. Ed. Norton Bryan G. Guildford, Surrey UK: Princeton University Press, 1986. 138–72. Print.

Carolyn Calloway-Thomas. *Empathy in the Global World: An Intercultural Perspective*. 2455 Teller Road, Thousand Oaks, California, 91320, United States: 2010. Print.

Charland, Louis C. "Emotion Experience and the Indeterminacy of Valence." *Emotion and consciousness* (2005): 231–254. Print.

---. "The Heat of Emotion: Valence and the Demarcation Problem." *Journal of consciousness studies* 12.8-10 (2005): 82–102. Print.

Cosmides, Leda, and John Tooby. "Evolutionary Psychology and the Emotions."

Handbook of emotions (2000): 91–115. Print.

---. "Evolutionary Psychology: A Primer." *Evolutionary Psychology: a primer* (1997).

Darwall, Stephen. "Empathy, Sympathy, Care." *Philosophical Studies* 89.2 (1998): 261–282. Print.

Dawkins, Richard. "The Selfish Gene." *revised edn. Oxford* (1989): Print.

De Waal, Frans. *The Age of Empathy: Nature's Lessons for a Kinder Society*. Random House LLC, 2010.

Dunbar, Robin, and Robin Ian MacDonald Dunbar. *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, 1998.

Gallese, Vittorio. "The Roots of Empathy: The Shared Manifold Hypothesis and the Neural Basis of Intersubjectivity." *Psychopathology* 36.4 (2002): 171–180. Print.

---. "The 'Shared Manifold' Hypothesis. From Mirror Neurons to Empathy." *Journal of consciousness studies* 8.5-7 (2001): 5–7. Print.

---. "The 'Shared Manifold' Hypothesis: From Mirror Neurons to Empathy." *Journal of consciousness studies* 8.5-7 (2001): 33–50. Print.

Gallese, V, C Keysers, and G Rizzolatti. "A Unifying View of the Basis of Social Cognition." *Trends in Cognitive Sciences* 8.9 (2004): 396–403. *CrossRef*. Web.

Goldman, Alvin I. "In Defense of the Simulation Theory." *Mind & Language* 7.1-2 (1992): 104–119. Print.

Gopnik, Alison, and Henry M. Wellman. "Why the Child's Theory of Mind Really Is a Theory." *Mind & Language* 7.1-2 (1992): 145–171. Print.

Greene, Sheila. "V. Biological Determinism: Persisting Problems for the Psychology of Women." *Feminism & Psychology* 14.3 (2004): 431–435. Print.

Currie, Gregory. "Empathy for Objects." *Empathy: Philosophical and psychological perspectives* (2011): 82.

Hoffman, Martin L. "Is Altruism Part of Human Nature?" *Journal of Personality and social Psychology* 40.1 (1981): 121. Print.

---. "The Contribution of Empathy to Justice and Moral Judgment." *Reaching out: Caring, altruism, and prosocial behavior* 7 (1987): 161–194. Print.

---. "The Contribution of Empathy to Justice and Moral Judgment." *Reaching out: Caring, altruism, and prosocial behavior* 7 (1994): 161–194. Print.

Kennett, Jeanette. “Autism, Empathy and Moral Agency.” *The Philosophical Quarterly* 52.208 (2002): 340–357. Print.

Krebs, Dennis L. “Altruism: An Examination of the Concept and a Review of the Literature.” *Psychological bulletin* 73.4 (1970): 258. Print.

Lawrence, E. J. et al. “Measuring Empathy: Reliability and Validity of the Empathy Quotient.” *Psychological Medicine* 34.5 (2004): 911–919. *CrossRef*. Web.

---. “Measuring Empathy: Reliability and Validity of the Empathy Quotient.” *Psychological Medicine* 34.5 (2004): 911–919. *CrossRef*. Web.

Liesen, Laurette T. “Women, Behavior, and Evolution: Understanding the Debate between Feminist Evolutionists and Evolutionary Psychologists.” *Politics and the Life Sciences* 26.1 (2007): 51–70. Print.

Lindgren, Henry Clay, and Jacqueline Robinson. “An Evaluation of Dymond’s Test of Insight and Empathy.” *Journal of consulting psychology* 17.3 (1953): 172. Print.

McClintock, Charles G. “Social motivation—A Set of Propositions.” *Behavioral Science* 17.5 (1972): 438–454. Print.

Newson, Lesley, and Peter J. Richerson. “The Evolution of Flexible Parenting.” *Evolution’s Empress: Darwinian Perspectives on the Nature of Women* (2013): 151–67. Print.

Nichols, Shaun. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford University Press, 2004.

Nichols, Shaun, and Stephen P. Stich. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds.* Clarendon Press/Oxford University Press, 2003.

Noddings, Nel. *Caring: A Relational Approach to Ethics and Moral Education*. University of California Press, 2003.

---. *Educating Moral People: A Caring Alternative to Character Education.* ERIC, 2002.

---. *Starting at Home: Caring and Social Policy*. University of California Press, 2002.

Pinker, Steven. *The Blank Slate: The Modern Denial of Human Nature*. Penguin, 2003.

Preston, Stephanie D., and Frans De Waal. “Empathy: Its Ultimate and Proximate Bases.” *Behavioral and brain sciences* 25.01 (2002): 1–20. Print.

Prinz, Jesse. “Against Empathy.” *The Southern Journal of Philosophy* 49 (2011): 214–233. *CrossRef*. Web.

Prinz, Jesse J. “Is Empathy Necessary for Morality?” *Empathy: Philosophical and psychological perspectives* (2011): 211–229. Print.

Richerson, Peter J., and Robert Boyd. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, 2008.

Rifkin, Jeremy. *The empathic civilization: The race to global consciousness in a world in crisis. Penguin, 2009.* Rizzolatti, Giacomo, Corrado Sinigaglia, and Frances Trans Anderson. *Mirrors in the Brain: How Our Minds Share Actions and Emotions.* Oxford University Press, 2008.

Sagi, Abraham, and Martin L. Hoffman. “Empathic Distress in the Newborn.” *Developmental Psychology* 12.2 (1976): 175. Print.

Slote, Michael. *The Ethics of Care and Empathy*. Routledge, 2007.

Sober, Elliott., and David Sloan. Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, Mass.: Harvard University Press, 1998. Print.

Stotland, Ezra. “Exploratory Investigations of Empathy.” *Advances in experimental social psychology* 4 (1969): 271–314. Print.

Stotz, Karola C., and Paul E. Griffiths. "Dancing in the Dark." *Evolutionary Psychology*. Springer US, 2003. 135-160. Tsoudis, Olga. “Influence of Empathy in Mock Jury Criminal Cases: Adding to the Affect Control Model, The.” *W. Criminology Rev.* 4 (2002): 55. Print.

Van Boven, Leaf, and George Loewenstein. "Empathy gaps in emotional perspective taking." *Other minds: How humans bridge the divide between self and others* (2005): 284-297.

Vischer, Robert. “On the Optical Sense of Form: A Contribution to Aesthetics.” *Empathy, form, and space: Problems in German aesthetics* 1893 (1873): 89–124. Print.

Waal, F. B. M. de. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, Mass.: Harvard University Press, 1996. Print.

Wilson, Edward O. *Consilience: The Unity of Knowledge*. Vol. 31. Random House LLC, 1999.

Zahavi, Dan. “Beyond Empathy. Phenomenological Approaches to Intersubjectivity.” *Journal of Consciousness Studies* 8.5-7 (2001): 5–7. Print.

---. “Empathy and Direct Social Perception: A Phenomenological Proposal.” *Review of Philosophy and Psychology* 2.3 (2011): 541–558. *CrossRef*. Web.

---. “Empathy and Mirroring: Husserl and Gallese.” *Life, Subjectivity & Art*. Springer, 2012. 217–254.

2. The care account and the role of values in empathy

2.1 Introduction

The main question in empathy research is "how does empathy work"? The majority of accounts that answer this question share three features. First, they focus on short-term phenomena "inside the head" of an individual agent. Second, the evidence they appeal to is gathered exclusively from laboratory observations. Third, they posit neurocognitive mechanisms that are (universally) shared by all humans. Accounts that have these three features can be said to be following the *conventional approach* to explaining how empathy works. A prominent account that follows the conventional approach is the *mirror-neuron account of empathy*. Mirror neurons fire both when we perform an action, and when we observe another organism performing a similar action. Their activation is recorded in hand-gesture experiments over short time periods in a laboratory setting, and is taken to fully explain how an agent empathizes with a target (Gallese, Keysers, and Rizzolatti 2004).

In this paper I develop an account of how empathy works that takes a different approach. I call this account the *care account of empathy*. The care account focuses on the value of care, and the role of values more generally in empathy. Values are important. They are at the core of an agent's relationship to the environment. For example, some basic emotions such as a fear, and anger are often said to provide an agent with evaluative relationships to the world that are fundamental to its survival (Panksepp 2001). Values have an affective, a cognitive, and a behavioural component (Rokeach, 1973). I follow Rokeach (1973) in taking a value to be "an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence." (Rokeach, 1973) Accordingly, values are evaluative beliefs insofar as they are about whether an end-state or behaviour is desirable or undesirable.

Conventional accounts of empathy have neglected values. At first glance, this is surprising, in part, because everyday goal-directed action is guided and motivated by

values. But ignoring the value of care is especially problematic for answering the question of how empathy works because agents that empathize accurately with a target feel motivated to help that target with their concerns. It is an agent's care or value placed on the well-being of a target that motivates that agent to help a target of empathy when it occurs. Furthermore, whether an agent cares for a potential target of empathy influences whether that agent attempts to empathize with that target. And whether and when an agent cares for a target is influenced by that agent's other values, which are differently operative in different environments. As psychological states then, the value of care and values generally should feature prominently in any account of empathy that minimally involves an agent sharing a target's psychological states.

This is not to say that, once empathy occurs, an agent will in fact help a target. For example, if an agent sees advertisements for charitable organizations asking for money to help starving children, that agent may empathize with those children by becoming aware of their concerns, feel negative emotions, and be motivated to help them. However, that the agent is motivated at a particular time to help these children does determine whether the agent will in fact help them. Other considerations (e.g. there being too many to help) may override this motivation or even eliminate it. Empathy merely requires that the agent be motivated to help the target at some point.

I distinguish between two types of empathy accounts: *matching-accounts* and *concern-accounts*. Then, I argue that care is a necessary component of the end state of empathy because it is what explains why agents feel motivated to help targets when empathy occurs. My account is a significant departure from all existing accounts of empathy that have not considered the roles of care in empathy. Furthermore, existing empathy accounts do not sufficiently take into consideration values more generally. My *care-account* fills this theoretical gap in empathy research. I outline the roles of care in the different component psychological processes of empathy, and model care in relation to how values influence when empathy occurs by shaping the motivations that agents have towards potential targets of empathy. I conclude that my care account provides a more precise explanation of empathy than other concern accounts and instances of the matching

account, and that it is better able to predict when empathy will occur. I will now begin by fixing reference on what I broadly take to be the phenomenon of empathy.

2.2 The phenomenon of empathy

Before I begin, I would like to make small point about imagination in reading and in empathy. Throughout this paper I will be presenting many examples that will allow me to fix reference on the phenomenon of empathy. These examples are drawn from my own experience, but I will intentionally present the persons in these examples using gender-neutral terms, and I will not describe their physical appearance. I have not done this so to sidestep the issue of gender parity in my research (or any other issue). Rather, I have done so for us to more easily notice the types of informationally-laden imaginations that we construct when reading. In reading this paper, you might, in a relevant sense, be attempting to take the perspective of the fictional targets in its examples. Similarly, when taking the perspective of non-fictional targets, we also construct and fill in certain information. This information is often based on our own experience and our own values. I think that presenting my examples in such neutral terms may allow us to further notice the importance of taking into consideration environmental variables and values in developing a fuller account of empathy. Now, on to a few examples.

Imagine that a friend tells you that they have recently been fired from a job they found fulfilling. Upon hearing this, you share their disappointment, their belief that they were dismissed unjustly, and you feel motivated to help them find a new job. You experience similar emotions and beliefs as they do, and you are motivated to help them appropriately. You empathize with your friend. It is also possible to empathize with someone you have not previously met. Imagine you are taking a leisurely evening walk around your neighborhood. You see a person sitting against lamppost bearing a cardboard sign asking for money. You imagine what it would be like for you to be in that person's place, you feel somewhat sad, and feel motivated to give them some change. You empathize with that stranger. It is also possible to empathize with an enemy. By imagining how an enemy perceives your actions and their effects as being dangerous or harmful, you can feel motivated either to change your actions or to better describe the

goals of your actions in order to help an enemy towards increased cooperation. These are examples that help us to understand *what empathy is*: becoming aware of someone else's concerns, sharing their emotions and beliefs, and feeling motivated to help them with their concerns. A point I will elaborate upon later is that empathy is often taken to be synonymous with mindreading. But there is a big difference between the two both in terms of their psychological processes and their neural instantiation (Singer 2006). Whereas *mindreading* involves understanding the behavioural intentions of a target, *empathy* invariably involves in some sense sharing a target's emotional states. While there is some debate about the neural overlap between the *mentalizing system* that makes mindreading possible and the *mirror (neuron) system* that is often said to be involved in empathy (Overwalle & Baetens, 2009, p 566), most theorists take the key component psychological processes of empathy to be *emotional contagion*, *mimicry*, *perspective-taking*, and *motivation*.

These component psychological processes of empathy are the focus of accounts that answer the question: how does empathy operate? The *conventional approach* is to answer this question by discussing the *mental format* of one or more of empathy's component processes. Accounts that answer the question in this way include: *simulationist*, *theory-theory*, and *direct-perception* accounts. On "*simulationist*" accounts for example, the processes of empathy are realized in mental *simulations* (Goldman 1992; Gallese and Goldman 1998; Goldman 2006; de Vignemont and Singer 2006). These simulations are taken to be *perceptual representations* that an agent creates when taking the perspective of a target. Similarly, "*theory-theory*" accounts answer the question of how empathy works by arguing that the processes of empathy are realized in mental representations that are less perceptual and more like a-modal *theories* (Stich and Nichols 1992; Nichols 2004). Both simulationist and theory-theory accounts espouse a metaphysical theory of mind according to which perspective-taking in empathy employs mental representations. Contrary to this view, *direct-perceptual* accounts of empathy answer the question of how empathy works without positing such a role for mental representations (Gallagher 2001, 2008; Zahavi 2001, 2008, 2011a, 2011b). On a direct-perception account, an agent has causally unmediated access to the psychological states of a target. An agent reads the

psychological states of a target off the target's behaviour. The target's states are thus directly available in the agent's conscious experience. What is important to note is that all three of these accounts (simulationist, theory-theory, direct-perceptual) answer the question of how empathy works by characterizing the mental format of the psychological processes of empathy.

But the question of how empathy works remains elusive. This is partly because the various accounts that focus on the mental format of empathy are actually talking about different phenomena. There is a striking variety of uses of the term 'empathy' among these accounts. And indeed there is no consensus in empathy research about what empathy is. This inconsistency of usage is impeding progress towards answering the question of how empathy works. By not clearly specifying what empathy is and focusing on the mental format of empathy, accounts that agree on the format of empathy turn out to disagree about what empathy is. With the aim of clarifying what empathy is, in the next section, I will discuss how four of empathy's component processes—(1) emotional contagion; (2) mimicry; (3) perspective-taking; (4) motivation—inform different uses of the term 'empathy'.

2.3 Various uses of 'empathy'

The term 'empathy' is applied to many different phenomena (Batson 2009). In this section I distinguish four phenomena that have been variously emphasized in connection with different uses of the term. My initial classification here follows from what most theorists take to be component processes that play important roles in empathy.

Process 1: Emotional Contagion

Emotional contagion is described as an agent involuntarily "catching" the emotions of a target (Doherty 1997). For example, psychological experiments show that 2-day-old infants cry when another newborn cries. This seems to be a specific response to the vocal properties of the other's cry (Simner 1971; Sagi & Hoffman 1976). Infants reacted in a more subdued manner to auditory cues of the same intensity that did not resemble human

crying (*Ibid.*). These and similar results have been taken to be evidence of a "rudimentary empathic distress reaction" (Sagi & Hoffman 1976). Sometimes emotional contagion is referred to as causing "affective empathy" whereby an agent experiences a vicarious emotional response to the emotions expressed by a target (Knafo et al. 2008 p 3; Zahn-Waxler et al. 1992). When 'empathy' is used to refer to emotional contagion, an empathic agent involuntarily matches the emotional state of a target.

Process 2: Mimicry (behavioural and neural)

Imitation of a target, or *mimicry*, is also sometimes called "empathy" (Bavelas et al. 1996; Hoffman 2000; Darwall 1998). Motor mimicry—which includes mimicking facial expressions, bodily movements and postures—has been proposed as an empathic response to a target. Neonates and infants can imitate both facial and manual gestures (Meltzoff and Moore 1977, 1983; Meltzoff 1988), and it has been shown that when adults are exposed to emotional facial expressions they often spontaneously mimic part of these facial stimuli (Dimberg et al. 2011). Likewise, adults also often mimic the postures and behaviours of others (Bavelas et al. 1986). And the feedback internal to an agent resulting from *motor mimicry* can initiate and modulate felt emotion (Darwall 1998, p 265; Ekman 1992; Adelmann and Zajonc 1989). Motor mimicry has been taken to be a "major empathic mechanism" whereby an agent matches the internal state of a target as a consequence of matching their behaviour (Darwall 1998, p 266). On such accounts, empathy ends when an agent vicariously experiences the emotional state of a target as a consequence of voluntarily or involuntarily mimicking the target's behaviour. Accordingly, even when an agent mimics a target's nervous tick, and experiences the same emotion as that target as a result of this mimicry, then empathy has occurred.

Neural Mimicry is another form of mimicry that has been the focus of explanations of empathy. Rather than focusing on mimicked motor behaviour (that may result in matching internal states), Preston & de Waal (2002) have proposed a "unified theory of empathy" according to which an agent mimics a target's neural state automatically and without mediation by behaviour. They argue that when an agent perceives a target, the agent mimics the target's neural representations because perception and action rely on the

same neural circuits. An agent produces an internal state that is partially the same as that of the target by neurally mimicking the target. Other influential accounts of empathy focus on another version of neural mimicry (which they call "mirroring") that involves *mirror neurons* (Gallese, Keysers, Rizzolatti 2004; Iacoboni 2008). Mirror neurons fire both when we perform an action, and when we observe another performing a similar action. On some uses of 'empathy', mirror neurons implement the processes by which we mimic and understand the internal states of others. It is important to note that there exist severe criticisms of the conclusions drawn from the empirical investigation of mirror neurons. For example, Hickok (2009) argues that we should not take mirror neurons to have semantic properties that would contribute to, if not fully explain, how we understand the goal of another's actions. Rather, he suggests that the role of mirror neurons may be to prime *motor vocabulary* by neurally instantiating associations between areas of the brain that respond to perceptual information (e.g. movement or sound) and areas of the brain that store action patterns (*Ibid.*). The implication here being that the language system plays a prominent role in understanding (mindreading) or empathy.[11] Similarly, Kilner and Lemon (2013) argue that the role of mirror neurons in contributing to understanding is at best incomplete. Like Hickok, they are critical of attributing semantic properties to mirror neurons. They cite evidence that top-down inputs from the visual systems and memory contribute to the generation of the models for predicting the sensory input that targets are providing (*Ibid.*).

Process 3: Perspective taking (like-me and like-other)

The third process which has been the focus of some uses of the term 'empathy' is *perspective taking* (Stotland 1969; Ickes 1993; Batson 1991; Ruby and Decety 2004). On this usage, what is important is that the agent imagine or infer what it is like to be in a target's place. Two modes of perspective have been distinguished. The first of these is the *imagine-self perspective* (Stotland 1969; Batson 1991). In this mode, an agent takes the

[11] Hickok (2009) presents other criticism along similar lines that mirror neurons should be seen as facilitating associations between perceptual information and other neural systems that process object recognition and semantic understanding such as that of the ventral stream of visual processing. He also points to a variety of dissociations between the mirror system and its purported psychological and behavioural effects that do not support its specified role.

perspective of a target by imagining what it would be like for the agent to be in the target's position. For example, if I as an agent take the imagine-self perspective of a target who is asking for money on a public street, I imagine what it would be like if I were in that target's situation. I will imagine what it would be like to be in the target's situation and compare that situation with my own current situation and past experiences. This results in my having a felt internal state that is similar to that of the target, and an awareness of some of the target's possible concerns.

The second mode of perspective taking is the *imagine-other perspective*. In this mode, the agent does not imagine what it would be like as that agent in the target's situation. Rather, the agent's perspective taking is based on an open sensitivity to the target's emotional states and on beliefs about the particular target–such as beliefs about the target's situation, concerns, and past experiences (Barrett-Lennard 1981). For example, suppose that I see a man asking for money on the street, and recognize him as someone I met last week. On that occasion, he and I interacted over the course of a minute or two and I came to believe that he is a military veteran who served his country in several tours of duty. This belief then plays a role in my imaginings of what it is like for him (as opposed to me) to be in his situation. The emphasis in imagine-other perspective taking is on what the particular target feels and is concerned with, rather than on what the agent would feel or be concerned with were the agent in the target's situation.

Process 4: Altruistic motivation

Some uses of 'empathy' take empathy to include an agent's motivation to help a target (Hoffman 1975; Batson 1991; Sober and Wilson 1998). For example, Batson's (1987, 1991) conception of empathy includes the production of what he calls *altruistic motivation* (*Ibid.*). His experiments show that when an agent empathizes with a target in distress, the agent will be motivated to help the target. He contrasts his use of the term 'empathy' with "personal distress". When an agent feels *personal distress* at witnessing the suffering of a target, an agent may choose to help the target as a way of relieving their own distress. Often, however, an agent will choose instead to relieve their own distress in a way that does not involve helping the target. Batson calls this "*egoistic motivation*".

Thus, on his use of the term, empathy includes an altruistic motivation that leads an agent to help a target in distress.

Hoffman (1975) also uses 'empathy' in a way that includes an altruistic motivational component. He cites the example (among others) of a child seeing another child with a cut finger. He argues that when the child "takes the role" or perspective of another child in pain, the former may experience "empathic distress" (Hoffman 1975, p 613). The agent then attributes this feeling of distress to the target. Consequently, the agent will be motivated to help the target (p 615). Like Batson, Hoffman also distinguishes between this form of *altruistic motivation* and more directly *egoistic motivation* (p 617). The tendency to altruistically help others in distress as a result of taking their perspective is central to Hoffman's account of empathy.

Some conceptions of empathy involve just one of these four processes. Others involve several of them. Take one of the reference fixing examples I presented at the outset: empathizing with a friend who recently lost their job. At least two of the component processes of empathy are involved. When my friend tells me about being dismissed, I may take their perspective by imagining how I would feel about losing my job and think about whether how I was treated fairly (perspective taking). I may also feel motivated to look for new employment in the same field or feel motivated to concentrate my efforts on finding different employment (motivation). The other component processes of empathy could also be involved. When my friend comes to me with the news of their dismissal, I may perceive what I take to be an expression of sadness on their face and vicariously feel sad myself (emotional contagion). Doing so, I may slouch in my chair, thus adopting a similar posture as my friend (behavioural mimicry). But the question arises, what counts as accurate (or successful) empathy in in this example? Is it that at least one of empathy's component processes is active? Or must more than one process be active for accurate empathy to occur? Having raised the question of what counts as accurate empathy, we can begin to see that answering the question of "how empathy works" by discussing only the nature (or format) of psychological processes will not suffice. The question is indeed premature because the question of what empathy is has not been sufficiently established.

We can better understand what empathy is by distinguishing between two types of empathy accounts: *matching accounts* and *concern accounts*. Each of these account types employs distinct criteria of empathic accuracy—the criteria for an empathizer to "get it right". Empathic accuracy is the result of empathizing; it is the end state of empathy. Accordingly, each account type's conception of what empathy is varies according to its criteria of empathic accuracy. I will now go on to describe this relationship between what empathy is and *empathic accuracy*.

2.4 The relationship between what empathy is and empathic accuracy

Empathy accounts differ in what they take empathy to be. And this affects when they take empathy to occur. Generally, when a person experiences empathy the process can be divided into three parts: 1) the moment it gets started; 2) the activity of processes that instantiate it; and 3) the moment it stops. Each account type of empathy differs with regard to the criteria that specify when empathy stops. And the output or state of an agent at that moment can be evaluated in terms of whether accurate or inaccurate empathy has occurred. These criteria for when empathy stops and an agent is truly empathizing (when an agent "gets it right") are criteria of *empathic accuracy*. There is a long tradition of explicit discussion of theories of empathic accuracy in the context of professional counselling and therapy (Fiedler 1950; Rogers 1957; Ickes et al. 1997). In this context, psychologists and psychiatrists have developed many empathic accuracy scales for measuring a therapist' (or counselor's) empathic accuracy when interacting with a client, and other empathic accuracy scales for various measurements such as "closeness" between married couples (Feldstein and Gladstein 1980; Ickes et al. 1997). It is important to note that all accounts of empathy have a theory of *empathic accuracy* (although it is often left implicit).

There is a significant conceptual problem in understanding the relationship between what empathy is and when accurate empathy occurs. The problem is that there is no principled relationship between which process an account focuses on and its theory of empathic accuracy. For example, an account may focus on neural mimicry as the main process of

empathy, and its theory of empathic accuracy may specify that an agent empathizes accurately with a target when the agent *matches* the target's internal state. Alternatively, an account may focus on perspective taking, and its theory of empathic accuracy may specify that an agent accurately empathizes with a target when the agent becomes aware of the target's needs or *concerns*. This problem is compounded when we notice that even accounts of empathy that focus on the same processes (e.g. perspective taking) can have different criteria of empathic accuracy. Any two accounts of empathy may focus on the same psychological processes of empathy but differ in terms of when those processes end and provide an end state to be evaluated in terms of accuracy. To take another example, Ickes's account of empathy focuses on perspective taking (Ickes 1993). And his theory of empathic accuracy states that empathy occurs when an agent accurately infers "the specific content of another's person's thoughts and feelings." (Ickes 1993, p 591) On Ickes's account, the criterion for accurate empathy is that the agent has *knowledge* of the target's thoughts and feelings (Ickes 1993, p 590-591). Compare this to Batson's account of empathy which also focuses on perspective taking. His theory of empathic accuracy requires that an agent experience an emotional state that is similar to that of the target.[12] The relationship between which processes an account empathy focuses on and its criteria for evaluating when accurate empathy occurs is problematically arbitrary because theorists have not agreed on what empathy is.

While some researchers have begun to discuss the conceptual difficulty resulting from the various uses of the term 'empathy' (and their associated theories of empathic accuracy) (Batson 2009), most seem to regard this divergent usage as merely terminological, and therefore unimportant. But indeed this variation is in fact of great importance because an account's theory of empathic accuracy will have consequences for what it counts as empathy. The question of *how* empathy works is premature because the question of *what* empathy is has not been satisfactorily answered (as is apparent in the diverse ways that empathic accuracy is used). As mentioned, I organize accounts of empathy into two types (*matching accounts* and *concern accounts*) according to their criteria of empathic

[12] I will describe Batson's account in more details after having presented my two-fold distinction between types of empathy accounts.

accuracy. Rather than attempting to identify what empathy is according to the different component processes that instantiate it, I tether what empathy is to two sets of criteria for what counts as accurate empathy. This allows me to consistently identify what empathy is across research that focuses mostly on its component processes. This way of specifying what empathy is will in turn allow me to return to the component psychological processes posited in different accounts and examine how they operate in relation to what counts as accurate empathy.

It may seem strange to speak of empathy in terms of a process with parts or stages because many philosophical analyses of phenomena take the form of specifying the *individually necessary* and *jointly sufficient conditions* for them to occur. But this shall not be the form of my analysis of empathy. I do not think that only providing a list of the conditions that specify when empathy occurs is the best way to understand it. My goal is not to provide a conceptual analysis of empathy whereby one uncovers the conception of empathy that all researchers share. Nor is it to provide a definition of empathy that will mitigate the inconsistent uses of the term. Researchers that provide an account of empathy do use the term consistently. Rather, my goal is to provide an analysis of empathy as a process and subsequently an account of empathy that takes into consideration certain neglected aspects of it such as values, environmental contexts, and motivations. I believe that analyzing empathy as a process will help us to better understand it by making explicit the *causal connections* between the antecedents and the effects of the psychological states it involves. If we only specify empathy as state, like Ickes (1993), whereby an agent infers "the specific content of another's person's thoughts and feelings", this does not tell us what causes an agent to attempt such an inference (Ickes p 591). Hence, it does not tell us when empathy starts. Similarly it does not tell us about the possible effects of empathy beyond the purported inference because there is no way of telling immediately when such an inference has occurred. But if we treat empathy as a process, we can better account for when it occurs and better explain the relationship between its causes and its effects. Let us take sex between humans as an analogical example. Sex can be treated as either a state with necessary and sufficient conditions or it can be treated as a process with a beginning, middle, and end. On the former analysis, we

may say for the state *sex* to occur it is sufficient that a penis is inserted into a vagina. On the latter analysis, we may say sex is the process of inserting a penis into a vagina. This analysis has certain advantages. We can take sex to be a goal-directed activity; we can explain it in terms of antecedent sexual arousal which often continues during the process of sex; and we understand the connection between its ending and its effects such as pregnancy. This is not the say that this is impossible when analyzing sex as a state. But merely specifying the conditions for the occurrence of state does not make explicit the relationship between the antecedent states that may be important components of the state or the relationship between its occurrence and its effects.

There are vast metaphysical assumptions that come with one's choice of analyzing a given phenomenon as either a state or a process.[13] But in the case of empathy as a construct, I believe that its use in scientific contexts (e.g. in empathy tests), lends some credence to treating it as a goal-directed process. Further, doing so will allow us to account for the causal connections that explain why an agent attempts to empathize with a target, and it will allow us to better predict when an agent is more or less likely to attempt to empathize with a target. Also, treating empathy as a process will be conducive to an evolutionary explanation that takes these explanatory and predictive goals into account. In my third paper, I show how understanding the causal connections between the psychological states involved in the process of empathy as it occurs in various environmental contexts allows us to better explain how it evolved.

2.5 Two types of empathy accounts

Most empathy researchers understand what empathy is by reference to the activity of the psychological processes it involves. As we have seen this leaves open the problematic possibility that even if two accounts are referring to the same processes, they may have different criteria of empathic accuracy. Rather than classifying accounts according to the processes they involve, I classify them according to their theory of empathic accuracy. That is, by the theory of when the process of empathizing has ended successfully. This

[13] See debates about *process metaphysics* and *substance metaphysics*.

allows me to differently sort accounts of empathy by what they take empathy to be. A possible worry at this point may be that one must know what empathy is before knowing what makes empathy accurate. But in my view, knowing what empathy is and knowing what makes empathy accurate come together. That is, they inform each other. Accordingly, I think it will be informative to treat empathy as a process and investigate its accuracy conditions in an attempt to better understand both what it is and its accuracy conditions. My choice of proceeding by investigating empathy's accuracy conditions is, in part, determined by what many researchers have said about the effects of empathy. Namely, that when empathy occurs, helping behaviour that is altruistically motivated often ensues. My investigation of empathy's accuracy conditions is an avenue into identifying what empathy is across differing accounts of the psychological processes it involves in virtue of the causal connections that my analysis reveals between the end state of empathy (empathic accuracy) and the activity of these processes.

2.5.1 Matching accounts

The main criterion of empathic accuracy on matching accounts is that there be a *match* between the internal state of an agent and the internal state of a target. There are two *matching relations* that matching accounts make use of. The first of these is an exact match between the thoughts and feelings of a target, and the *"inference"* or *representation* that an agent has about the target's thoughts and feelings.

> **Matching relation A:** An agent *infers* (or represents) the content of a target's experience.

Mentioned above, Ickes's (1993) account is a token of a matching account. The relevant match on this account is between an agent's represented inference and a target's experience. The procedures he and his colleagues have developed to assess empathic accuracy is called the "dyadic interaction paradigm" (Ickes et al., 1986; Ickes et al., 1990a, 1990b; Ickes and Stinson, 1992). It involves two participants interacting for five minutes while unknowingly being videotaped. The participants are then informed that they were being videotaped and are then separated. Once separated, the participants are asked to watch the video of their interaction and to provide a written record of (a) their

own thoughts and feelings during the interaction, and (b) what they believe their interaction partner was thinking and feeling during the interaction. On the first viewing, the participants are instructed to pause the video at each point at which they remember having a specific thought or feeling. They are then asked to watch the video a second time and report what they believed their partner was thinking and feeling at each point that the video was paused during the first viewing. The video is then watched a third time, and each participant is asked to report on the thoughts and feelings of their interaction partner at every time the video was paused during the first and second viewing. Ickes's account is a matching account because empathic accuracy occurs when there is an exact match between what an agent represents a target's experiences to be and that target's actual experience.

Jesse Prinz also argues in favour of a matching account of empathy. On Prinz's account, accurate empathy occurs when an agent's felt emotional state matches the felt emotional state of a target.

Matching relation B: An agent has the same experiences those of a target.

Prinz's is a matching account because the primary criterion of accurate empathy is that of *emotional experience* matching between agent and target. As he puts it: "In empathy, we feel the same emotion that someone else is feeling; we put ourselves in another person's shoes." (Prinz, 2007, p 82) He states that empathic responses can be measured by comparing the brain activation associated with the felt emotions of agents and targets (*Ibid.*). The matching relation on Prinz's account is different from that of Ickes. For Ickes, accurate empathy occurs when an agent's representations match the experienced states of a target. For example, if an agent accurately assesses that a target is sad, then that counts as accurate empathy even if the agent is gleeful about it. Whereas for Prinz, accurate empathy occurs when an agent experiences emotional states that are the same as (or match) those of a target's emotional states (Prinz 2011a; Prinz 2011b).

2.5.2 Valence

The concept of valence is important to the second type of empathy account (concern accounts). For this reason, I will describe it here before proceeding. Imagine a longtime salesperson in a large downtown retail store who has been working towards being promoted to manager of the store for several years. One day, this person is informed that the current store manager will be retiring next month and that they will likely be offered the position. Getting the new job would mean a substantial increase in pay and additional health benefits. A month goes by. But then the employee is informed that due to a downturn in sales projections, a new manager with more experience and an educational background in marketing will be hired instead.

We can easily imagine a significant difference between the emotions of the agent before and after being informed that they would not be getting the job. The emotional experiences that the employee has after being informed that the position is opening up and prior to being informed that they will not be offered the position are likely *positive*. After being informed that they will not be getting the job the employee's emotional experiences are likely *negative*. The distinction between positive and negative emotional experience is called *affect valence* (Charland 2005a; Colombetti 2005).

As applied to emotional experiences, affect valence refers to the agent's evaluative appraisal of an experience (Lazarus, 1991). On a subtly different use of the term, emotions themselves are taken to be either positive or negative (Prinz, 2004). For example, Prinz claims that some emotions, like anger, are 'negative' emotions. When the concept of valence is applied to an emotion in the abstract, rather than to the subjective evaluation of a particular felt experience by an agent, this is called *emotion valence*. It is important to note that *affect valence* and *emotion valence* are distinct because some emotion theorists believe that certain emotions are intrinsically valenced (Prinz 2004); whereas other theorists believe that the same emotion, say sadness, can be evaluated as having a different valence at different times (Charland 2005b).[14] For present purposes, I

[14] In the next paper I discuss the issue of intrinsic valence in relation to the role of unconscious emotions in accounts of empathy.

will take ‘valence’ to mean ‘affect valence’ (the valence of a felt emotional experience). Valence plays an important role on some accounts of empathy; what I call *concern accounts*. Let us now proceed to this second account type.

2.5.3 Concern accounts

The second type of empathy account is the *concern account* (Batson 1987, 1991; Hoffman 1975, 2000). What concern accounts and matching accounts have in common is that *matching* is a criterion for empathic accuracy. For example, on Batson’s account accurate empathy requires that the *valence* of an agent’s emotional experience be “congruent” with the *valence* of a target’s emotional experience (Batson 1991). By ‘congruence’, Batson does not mean that the content of an agent’s emotional state be the same as that of a target’s. For example, when an agent empathizes with a target, the agent may feel sad and the target may feel frustrated. Rather, what the agent must match is the *valence* of the target’s emotional state–the agent must experience a negatively valenced emotional state when a target is experiencing a negatively valenced emotional state (and likewise for positive valence).

> **Matching relation C:** The valences of the agent’s experience match the valences of the target’s experience.

But although matching is a criterion of empathic accuracy on both types of accounts, matching is *not* the main criterion on *concern accounts*.

On concern accounts, empathy is not just a matter of matching (Batson 1991; Hoffman 2000). Concern accounts add two criteria of empathic accuracy other than matching. The first of these is that an agent must become *aware* of the needs or concerns of a target. To achieve such awareness, the agent must have the ability to recognize that the target is an animate organism that is distinct from itself, other animate organisms, and inanimate organisms (Hoffman 1975, 2000). Then by taking the perspective of a target, the agent may come to have an awareness of the target’s concerns. On concern accounts, this is when empathy starts. And an accurate awareness of a target’s needs is a main criterion for assessing when accurate empathy has occurred.

The second additional criterion for empathic accuracy on concern accounts is that an agent must feel motivated to help a target with their concerns. This motivational component of empathy is supposed to be caused by the agent having matched the emotional valences of a target's internal states. An agent must select a target, take their perspective, and become aware of the target's concerns. This awareness of the target's concerns can result in the agent having an emotional response whose valence matches that of a target. Alternatively, the valence match between an agent and target can occur via processes that are less voluntarily controlled than perspective taking, such as emotional contagion. But regardless of the process by which an agent comes to have an experience whose valence matches that of a target, once the agent has such an experience, the agent must also feel motivated to help the target with their concerns (Hoffman 1981, p 51). And these concerns may be associated with either positively or negatively valenced emotional states. For example, a target may have just won a race and is expressing positively valenced joy. On a concern account, it is not sufficient that the agent vicariously experience the target's joy for accurate empathy to occur. The agent must experience joy or some other positively congruent emotional feeling such as pride. But it is also necessary that the agent become aware of the target's concern(s). An example of a target's concern in such a situation may be that others express recognition of how difficult it was to win the race. Thus, for accurate empathy to occur on a concern account, the agent must be aware of this concern for others to express recognition, and the agent must feel motivated to recognize the winning racer's achievement. It is important to note that behaviour directed at helping a target with their concern does not need to occur for accurate empathy on concern accounts. The agent must feel motivated to help, but does not need to act on this motivation. Helping behaviour may be suppressed. Or once motivated to help the target with their concern, the agent may be distracted or decide to engage in some other activity. However, the two main criteria for accurate empathy that concern accounts add are that an agent must first become aware of a target's concerns, and second, feel at least momentarily motivated to help that target.

To recap, concern accounts and matching accounts of empathy both specify that an important criterion for accurate empathy is that agents must (in some sense) match the

internal state of a target. On matching accounts this takes the form of an agent matching the experience of a target, or of an agent having matching representations about the target's experience. On concern accounts, the matching relation between agent and target is one of valence. For empathy to occur, an agent does not need to match the exact content of a target's emotional experience. But the agent must have an emotional experience that matches a target's experience in terms of its valence (either positive or negative). What differentiates *matching accounts* from *concern accounts* then is that the latter introduce two additional criteria for empathic accuracy: 1) that an agent have an experiential awareness (based on the target's emotional state valences) of the concerns of a target; and 2) that an agent feels motivated to help the target with those concerns.

2.6 Empathy as a motive

Many theorists unequivocally take empathy to motivate helping behaviour (Batson 1991; Sober and Wilson 1998; Hoffman 2000; Preston and de Waal 2002; Decety and Jackson 2004; Slovic 2007). This is especially true of those who put the concept of empathy to practical use, such as evaluating the relationship between therapists and patients (Dymond 1948; Fiedler 1950; Hall and Bernieri 2001). And the motivational contribution of empathy has been observed in everyday cooperative social life (Main 1979; Sawin 1979). Feeling empathy for a target motivates an agent to help that target with their concerns such as relieving their suffering or recognizing their achievements. Additionally, compelling evidence that empathy motivates helping behaviour is found in the behaviour of people who have difficulty empathizing with targets, such as people who meet psychopathic or sociopathic criteria. People meeting those criteria are capable of reasoning, making predictions about their own and others' behaviour, understanding principled norms, and experiencing emotions; indeed they often appear to be quite socially adept. But they either suppress their empathic motivation or lack the capacity for empathy altogether. It is their lack of empathic development that that explains their frequent disregard for the well-being of others (Anderson et al. 1999; Damasio 1994; Blair 1995, 2004, 2007). Concern accounts of empathy incorporate the motivational component of empathy better than matching accounts. But neither account type

sufficiently accounts for how values interact with other components of empathy to significantly affect how empathy motivates helping behaviour. Before addressing this issue, I will discuss how matching accounts and concern accounts treat empathy as a motive.

Matching accounts downplay empathy's motivational component. Rather, their focus is on the accuracy of the match between the internal state of an agent and that of a target. They do not discuss the motivational states involved in such matching. For example, an agent may need to be in a certain *motivational state* in order to take a target's perspective. Another example of how motivational states influence empathy is when an agent experiences a motivation to switch between modes of perspective taking (*like-me perspective taking* to *like-other perspective taking*) after matching has occurred. In later sections, I will expand on the roles of motivational states in empathy. For now, I am merely pointing out that matching accounts do not model an agent's motivation to help a target with their concerns while empathic matching is occurring. Matching accounts emphasize the matching relationship between agent and target rather than the motivational component of empathy that often leads to continued accurate empathy and helping behaviour.

On the other hand, concern accounts of empathy better explain the motivational component of empathy. Concern accounts tend to focus on the agent's awareness of a target's concerns. In doing so, they model how an agent achieves this awareness, and the relationship between empathy and helping behaviour. For example, Batson and his colleagues have performed many and various experiments demonstrating the factors that influence when an agent empathizes with a target and subsequently helps that target (Batson 1987, 1991; Batson and Shaw 1991; Batson et al. 1997; Batson et al. 2007). The overarching research goal of these experiments is to determine whether an agent is helping a target as a consequence of *egoistic motivation* or *altruistic motivation*. To this end Batson et al. isolate and manipulate variables that affect the relationship between empathy and helping behaviour such as:

i) Different modes of perspective taking
ii) Facial expressions
iii) Perceived similarity to a target
iv) Options for an agent to reduce their own distress at the perception of a target in need rather than helping that target
v) Punishment administered to an agent for not helping (e.g. negative social evaluation, negative self-evaluation)
vi) Reward administered to an agent for helping
vii) Having a "mood-enhancing" experience as a result of helping [15]

What Batson et al.'s experiments show is that there is a strong correlation between an agent empathizing with a target and an agent subsequently helping that target from *altruistic* motivations rather than *egoistic* motivations–that is, rather than being motivated to help for the advantages or disadvantages that are expected to result from helping or not helping. Including motivations as a component of empathy is very important. Let us return to an example above of a potential target of empathy winning a race and feeling pleasure. An agent might infer that the target who just won the race is elated (positive valence) and might, via perspective-taking, feel delighted (delighted). But the agent may not necessarily feel motivated to help that target. Although the agent and the target share the relevant emotional experience, the agent may in fact be matching the target with the intent of causing the target harm by reporting them for the use of disqualifying substances. We would not want to say that the agent in this case is accurately empathizing with the target precisely because the agent feels motivated to harm rather than help the target. Being able to distinguish between such cases is a very important reason for including the motivational component in an account of empathy.

But although concern accounts of empathy incorporate the motivational component of empathy better than matching accounts, neither account type specifies the role of *values* that affect this motivational component. In the next section, I will argue that the value an agent places on the well-being of target affects empathy in four different ways. I call this

[15] For a detailed summary of Batson et al.'s experiments on the relationship between empathy and altruistic helping behaviour, see Batson 1991, Part III.

value *care*. I will show that current accounts overlook the many roles of care in empathy. Considering the roles of care, as a value, in empathy opens up a space for presenting my *care account*. And in the remainder of the chapter, I will develop a sketch of my care account of empathy by connecting the role of care to that of values more broadly, and their influence on other components of empathy (such as goals and motivations) that existing accounts have neglected.

2.7 Care in empathy

In this section I will show that care—the *value* that an agent places on the well-being of a target—plays a central role in empathy. First, it should be noted that the term 'value' is used in several ways. In this section I will be using 'value' in reference to a value that an agent holds. I take value, on this usage, to be a subjective state of an individual agent. The agent "values" the well-being of a target. I will not be characterizing this value as being derived from a theory specifying when one target is objectively more valuable than another. Rather, the value that an agent assigns to the well-being of a target is a personal state of an agent that is derived from that agent's experience (or development) and is dependent on the agent's social context for its realization. Later, I will present and argue in favour a new account I call the *care account of empathy*. I will show how the value assigned to the well-being of a target, or *care* that an agent has for a target, varies according to which component processes of empathy are active. I will also show how care for a target varies according to specific motivational and social variables that importantly affect when empathy gets started, how it occurs, and under what conditions it is taken to be accurate. This will have significant implications for our understanding of what empathy is, and in turn, how it works.

The *value* that an agent assigns to the well-being of a target is what I call *care*. My notion of care as a value finds some support in the work of philosophers that are interested in what makes people able to achieve and maintain a state of care for a target. For example, Noddings (2003) states that care is a "capacity for interpersonal attention." (Noddings 2003, p 19) Her focus is on characterizing the practice of care and on arguing for why it is important. Similarly, Gilligan (1993) discusses the differences between men's and

woman's psychological and social development of an "ability to care." (Gilligan 1993, p 17) Like Noddings, Gilligan focuses on the expression of care. She argues that woman have historically taken on more caring roles than men, and that this contributes to the perpetuation of norms that contribute to specifying contemporary genders. I agree with Gilligan and Noddings that it is important to examine the development and expression of our capacity to care, especially when it is mediated by empathy. Agents who care for the well-being of targets have the capacity to do so, and express care in empathy. But for the purposes of this paper, I will not be focusing on the historical or evolutionary reasons why agents come to care about targets. Rather, I treat care as a value in order to investigate it as part of the broader set of values that agents have. Doing so allows me to describe the roles of care in empathy while clarifying its relationship to other values.

One of care's roles in empathy is that of valuing the well-being of a target which contributes to empathy "getting started". Along these lines, Batson has recently argued that "valuing the other's welfare" is a "key antecedent" to becoming aware of a target's concerns (Batson 2011, p 41). Here Batson means that care, as a "key antecedent", regularly causes some of empathy's component psychological processes to occur (e.g. perspective-taking). In the sections immediately below, I will be examining in detail the connection between care and empathy's psychological processes. Another of care's roles in empathy is that once the process of empathy is started it can cause care–it can cause an agent to value the well-being of a target.[16] This in turn could lead to further empathizing or more accurate empathy. I shall address these two roles in turn. In later sections I will describe how other values that both agents and targets have can also affect when empathy gets started and empathic accuracy.

2.8 The causal roles of care

The central claim of the care account of empathy that I will present is that an agent's *care for* a target is an individual value that is constitutive of empathy's end state. When an agent accurately empathizes with a target, that agent cares for the well-being of the target. This is indicated by an agent's motivation to help a target when empathy occurs. In other

[16] In this role care is a consequence of empathy rather than a cause.

words, the care in empathy explains why an agent feels motivated to help a target. Concern accounts are better poised to integrate the role of care in empathy than are matching accounts. Existing concern accounts suggest that care is involved by including a motivational component in their understanding of what empathy is (Stotland 1969; Hoffman 1990; Batson 2011). But the status of care and its roles in empathy have not been clarified.

I will describe four roles of care in empathy:

1) **Care: before *involuntary* perspective-taking**
2) **Care: after *involuntary* perspective-taking**
3) **Care: before *voluntary* perspective-taking**
4) **Care: after *voluntary* perspective-taking**

Then I will argue that in all of these instances, care for the well-being of a target must be present at the *end state* of empathy. Hence, that care is *constitutive* of empathy.

2.8.1 Care: before *involuntary* perspective-taking

Some empathy researchers claim that there are "two routes", internal to an agent, which lead to empathy: the involuntary and the voluntary.[17] This is a claim that I find well empirically supported, and that I agree with. Empathy can result from processes that are initiated involuntarily. On this involuntary route to empathy, *care* for the well-being of a target can be prior to the component processes of empathy becoming active.

When an agent perceives a target, psychological processes that may be activated involuntarily include *emotional contagion*, *motor mimicry*, and *neural mimicry*. Via the involuntary activation of these processes, an agent takes at least part of the perspective of a target. For example, the agent may experience an emotion that is exactly the same as that of the target; or the agent may experience an emotion that is congruently valenced with that of the target; or the agent may come to some awareness of the target's concerns

[17] Goldman (2011) calls these the "mirroring route" and the "reconstructive route".

and so on. But it is important to note that these involuntarily processes leading to perspective taking are not always activated upon the perception of all targets. When an agent perceives a target that the agent at least minimally cares for, some combination of emotional contagion, motor mimicry, or neural mimicry may occur. And fortunately, most people have at least a minimum of care for the well-being of targets that they know and like, and even targets that they have never met before. Accordingly, on the involuntary route to empathy, care is often a precondition for empathy to occur. Care can be an antecedent to perspective-taking that facilitates its activation.

This is not to say that care is a sufficient condition for empathy to get started before *involuntary* perspective taking. Although an agent may minimally care for a potential target of empathy, an agent may have difficulties with the involuntary activation of the component processes (e.g. neural mimicry) leading to perspective taking. As mentioned, persons meeting the criteria for sociopathy or psychopathy frequently disregard the concerns of targets. This may be the result of problems with the activation of involuntary empathic processes. This does not imply that such people do not necessarily care about a target. Even though they may care for a target, they may simply not be able to *involuntarily* take that target's perspective.

2.8.2 Care: after *involuntary* perspective-taking

Care also plays a role in empathy after *involuntary* perspective-taking has occurred. Independent of an agent having a disposition to care for a potential target of empathy, when involuntary processes lead to perspective taking, care for that target *can* ensue. Once an agent has taken the perspective of a target via the involuntary route to empathy, the agent may become aware of (some of) that target's concerns, and experience an emotional state that is *congruently valenced* with that of the target. When accurate empathy occurs, this perspective-taking will in turn motivate the agent to help the target with their concerns. That empathy motivates helping behaviour indicates that empathy can contribute to an agent valuing the well-being of a target. However, as Batson's experiments show, that an agent feels motivated to help a target (or even exhibits helping behaviour) does not necessarily imply that the agent is motivated by empathic

perspective-taking at all (or alone). There are many routes to helping behaviour that do not involve empathy. I will soon describe some of these routes in the section on care as subsequent to *voluntary* perspective taking. For now, suffice it to say that involuntary perspective taking can lead to care (as evinced by an agent's motivation to help a target with their concerns).

2.8.3 Care: before *voluntary* perspective-taking

That an agent care for a target is *not* required for voluntary perspective taking. I will discuss the notion of *voluntary perspective-taking* in detail in subsequent sections. A distinction between two types of perspective-taking significantly informs current debates and impacts experimental designs in empathy research. For present purposes, we can understand voluntary perspective-taking as an act that an agent performs whereby an agent imagines how a target's situation is affecting that target, and what that target's concerns are with regard to their well-being and the ongoing development of that situation (for better or worse). An agent can voluntarily take the perspective of a target without first caring for that target. And often in empathy experiments, agents are instructed to take the perspective of targets that they do not already care for (Stotland 1969; Batson 1991; Ickes 1993). However, if an agent cares for a target, that agent is more likely to take the target's perspective. Contrarily, an agent who places *no value* on the well-being of a target is not likely to imagine what it is like to be in a target's situation and how they are affected by it. Similarly, an agent who places value on harming a target is likely to experience incongruent emotional states to that of the perceived target in need.[18] For example, a salesperson may experience joy upon hearing that a competitor has not secured a potentially lucrative contract.[19] Or an agent may dispassionately understand a target's concerns without experiencing a congruent emotional state or feeling motivated to help (Batson 2011, p 41). However, when an agent cares for a target, this contributes to the likelihood of that agent taking the target's

[18] Unless the agent is attempting to deceive the target as in the example of the race winner above.
[19] I will discuss the relationship between various "*motivational orientations"* (e.g. competition) and values in empathy later in the paper.

perspective and in turn empathizing with that target. Care for a target is not required for an agent to start empathizing with a target, but it helps.

2.8.4 Care: after *voluntary* perspective-taking

As a result of perspective-taking an agent can come to care for a target, even if the agent did not care for the target prior to perspective taking. Even if at first an agent does not care for a target, once an agent has taken the perspective of a target, the agent may become aware of a target's concerns, and feel motivated to help the target with those concerns. Here again, that empathizing motivates an agent to help a target provides some reason to believe that the agent *cares for* the target.

This claim is empirically supported by the results of perspective taking experiments. Mentioned above, Batson et al.'s experiments isolate many of the factors that contribute to an agent being *egoistically motivated* (as opposed to *altruistically motivated*) to help a target. What their experimental manipulations show is that, when empathy occurs, an agent is primarily motivated by an awareness of the target's concerns as opposed to being primarily motivated by other factors such as perceived similarity, reward, punishment and so on. While these latter factors may exert influence on an agent's motivation to help a target, I think Batson et al.'s results also show that care and perspective-taking often contribute to an agent's motivation to help. As mentioned, Batson has recently argued that an agent needs to care about whether a target has concerns and about how those concerns affect the target (Batson 2011, p 41). Furthermore, several additional studies show that by age 3-5, children recognize the feelings of others (Borke 1971; Feshbach and Roe 1968). In one of these studies, children aged 3-8 years old were shown drawings of faces taken to represent happy, sad, afraid, and angry emotional states and asked to identify them as such (Borke 1971, p 264). If a child had difficulty identifying them correctly, the examiner identified the emotion for them. The children were then told stories in which another child would likely experience one of these emotions (e.g. losing a toy; getting lost in the woods at night). The stories were accompanied by a picture of a child with a blank face engaged in the described activity. Each participant child was then asked to select the face that best showed how the child in the story and associated picture

felt. In this same study, children were then shown eight additional stories in which they were described as behaving toward another child in ways that might make that other child happy, sad or angry (e.g. pushing the other child off a bike, sharing candy). In this second task, each child was asked to point to the face (happy, sad, or angry) that best indicated how the other child felt in the situation described. The results showed that approximately 88% of children were able to associate the appropriate face to the picture in the first task by age 5. The results were similar on the second task when the child had to identify the emotional state of a target they were interacting with. The children's ability to recognize the feelings of others increased with age. And, when agents do so, they often report both feeling emotions that are congruent with that of targets and being motivated to respond with helping behaviour.[20]

There are many variables (e.g. potential reward or punishment) that influence whether an agent's perspective-taking results in a motivation to help a target with their concerns. Helping behaviour often occurs in situations where several incentives have influenced an agent's motivation to help a target. However, the value that an agent assigns to the well-being of a target is often that agent's primary motivation to help. When care accompanies perspective-taking, empathy can occur.

2.9 On the separation of care and empathy

I have outlined roles that care plays in influencing when empathy gets started and how care in empathy motivates helping behaviour. In the next section I will make the stronger claim that care is *constitutive* of empathy. I will do so by examining whether the component psychological processes of empathy *typically* involve care or *require* care for their activation. But first, I will address a possible objection to my claim that integrating the role of care into an account of empathy is important. Specifically, I will address Prinz's (2011a; 2011b) claim that care and empathy should be analyzed or treated separately.

[20] For a review of 20th century studies pertaining to the relationship between perspective taking and helping behaviour, see Hoffman 1982.

I have used "care" to refer to the value that an agent assigns to the well-being of a target.[21] It is important to note that an agent can care for a target without empathizing with that target at all. As shown above, an agent can care for a target prior to taking that target's perspective. Or an agent can assign value to the well-being of a target according to more abstract principles that the agent takes to be contextually appropriate. Care can occur independently of empathy. For example, a potential target may not care for their own well-being and feel sad about this, but an agent can still assign value to the well-being of this target without taking that the target's perspective. But although it is possible to *care for* a target in this way, I have argued that care for a target *can* (and often does) result from perspective-taking. For example, a target who has recently lost their job may be concerned about their well-being and be concerned about finding a new job. An agent can come to care for this target by taking that target's perspective, sharing their emotional states, and feeling motivated to help them with their concerns. Similarly I will argue that care for a target can occur as result of perspective-taking even when the target does not care for themselves. For now, suffice it to say that these examples begin to suggest that care may play an important role in empathy.

Contrary to this, Jesse Prinz notably maintains that *care* and *empathy* should be treated as separate phenomena in all cases (Prinz 2011a, 2011b). Prinz's claim has both ontological and methodological implications. To assess it, we must first understand his distinction between care and empathy. For Prinz care is "worrying about [a target's] welfare" (Prinz, 2011a, p 211). He goes on to say that it is something one can do "even if one doesn't feel what it would be like to be in [the target's] place." (*Ibid.*) It is important to note here to for Prinz, concern (as opposed to my notion of care) is characterized by the emotion of worry, and that it occurs independently from an agent taking a target's subjective perspective. That is, being concerned for a target is a phenomenon that occurs from a third-person perspective (we may call this *abstract care*), separate from empathy. Recall that empathy for Prinz involves an agent matching in experience the psychological states of a target. It is a first-person response. Accordingly, Prinz thinks that it should be treated separately from concern (a third-person response). Prinz and I agree that care is a value

[21] In the history of philosophy, this has sometimes been called "sympathy" (Hume 1739/1978).

that an agent assigns to the well-being of a target. But he also thinks that care *does not* result from emotional matching between an agent and a target. That is, care does not result from empathy. Prinz provides two reasons in support of treating care and empathy separately.

Prinz uses the words "care" and "concern" interchangeably. In both cases it is the agent that is caring for (or is concerned for) a target. For simplicity, I will refer to this as "*Prinz's notion of concern*" in order to distinguish it from my notion of "*care*". So, the first reason Prinz provides for treating care and empathy separately is that his notion of concern is a *third-person* response. On this notion of concern, an agent worries about a target by having beliefs about what a target *should be* experiencing, as opposed to an awareness of what that target *is* experiencing. He provides the example of meeting a cult member who is delighted by their cult leader's nefarious plans (Prinz 2011a, p 2). When meeting the cult member, Prinz says, an agent will likely feel fear (of the leader) on the cult member's behalf because that agent knows that this is what the agent *should* feel (*Ibid.*). In this way, the agent is *concerned* for the target. On the other hand, empathy for Prinz is a first-person response. It is a response which allows an agent to experience the effects of a situation from the target's first-person point of view. Returning to the example, when an agent *empathizes* with the cult member who is delighted by their cult leader, the agent will also feel delighted. This accords with Prinz's account of empathy (which is a matching account) on which accurate empathy occurs when an agent comes to have the same psychological states as a target. Thus, although Prinz's notion of *concern* (like my notion of care) implies that an agent assigns value to the well-being of a target, he thinks that this value (expressed as worry) should be treated separately from empathy. This is because concern is a third-person response, and empathy is a first-person response.

The second reason Prinz provides to support the separation of concern and empathy is that concern (unlike empathy) is always an agent's *negatively valenced* emotional state (regardless of the target's emotional state valences). This allows Prinz to maintain that an agent can be *concerned* for targets that are plausibly emotionally quite different from that agent, or even for inanimate objects. The examples he provides are of being concerned

for a building that is in disrepair, a plant, or an insect (*Ibid.*). As mentioned, Prinz's notion of concern characterizes an agent's belief about the well-being of a target being in jeopardy and an agent's negatively valenced feeling of worry. It is not an emotional state that needs to be shared between agent and target. An agent can feel concern for a target, even though the target may not feel concern for themselves. The example he provides here is that of an agent seeing a drug addict using a drug. Like the agent who is worried about the cult member, this example of the drug addict is one in which the valence of the agent's states do not match those of the target. But this latter example adds a further important consideration: that the target does not care for themselves. Accordingly, the agent may be *concerned* for the well-being of the drug addict, even though the agent does not match the emotional state of the drug addict or the lack of concern that the target has for themselves. Whereas the emotion the drug addict is experiencing is pleasure, the emotion of the target is concern (Prinz 2011b, p 230). The target does not care for themselves, but the agent nonetheless values the well-being for the target (*Ibid.*). The agent can be concerned for the target without empathizing with the target. And vice versa, the agent can empathize with the target without being concerned for the target.

This is not the notion of empathy that I espouse. Unlike Prinz, I include a motivational component in my account of empathy. On my care account, like on other concern accounts, empathy is achieved when an agent becomes aware of a target's concerns and feels motivated to help with those concerns. I believe that this motivational component should be included because of the evidence that empathy, even on a matching account, often causes an agent to care for a target. On a matching account, there are counter-examples to this evidence. For example, an agent can match the emotional state (e.g. joy) of a target that has just won a baseball game. But the agent in this case will not be motivated to help the target with their concern for celebrating this win. This is because the agent in this example favoured a player on the opposing team to win. My care account of empathy treats such cases as attempts at empathy that end unsuccessfully. It is because the process of empathy has started in this agent that the agent matches the target's joy. But it will turn out that the agent will not be motivated to celebrate with the target because the agent is in a *competitive motivational orientation* towards the target

that contributes to this motivation being absent or changed. I will say more about the role of motivational orientations in later sections.

Care plays an important role in affecting the efficiency and accuracy of attempts at empathy. Specifically, if an agent cares for a target prior to selecting that target as a potential target to empathize with (either as a result of abstract care or previous empathizing), then it is more likely that the agent will succeed in empathizing accurately with that target. This is because for an agent to accurately empathize with a target, that agent may be required to voluntarily sustain an effort of perspective-taking leading to an awareness of the target's concerns. If an agent cares for a target prior to selecting that target as a potential target to empathize with, then I believe it will be more likely that this agent will feel motivated to help the target with their concerns. This causal relationship between the motivational component of empathy and helping behaviour is supported by evidence from evolutionary psychology presented by Baron-Cohen (2003) and Pinker (2011).[22]

The second important difference between my account of empathy that of Prinz' is that the notion of *care* that I employ is significantly different from Prinz's notion of *concern*. *Prinz's notion of concern* is more limited than my notion of *care*. For Prinz, concern is (1) third-person response, and (2) is always negatively valenced. I will address the second of these properties first.

I incorporate a more flexible assignment of value to the well-being of targets than Prinz's notion of *concern*. As we have seen, perspective-taking can result in a valence match between agent and target. Thus, if a target is in a positively valenced emotional state, an agent can take that target's perspective, become aware of their concerns, and feel a positively valenced care for that target. Care, unlike Prinz's notion of concern, is not always negatively valenced. On my account, an agent's *care for* a target can be experienced either as a negative or positive emotional state. By widening the notion of care for the well-being of a target to include the possibility that an agent can express care that is both positively and negatively valenced, my account characterizes an agent

[22] I address evolutionary psychological accounts of empathy in detail in paper 3.

assigning value to the well-being of a target in a wider variety of cases—namely, in cases where an agent is empathizing with a target in a positively valenced emotional state. It is important for an account of empathy to be able to handle such cases because agents often empathize with targets that have concerns while experiencing positively valenced emotions. For example, when a runner wins a race, that runner may be concerned with celebrating this achievement with others.[23]

The other relevant feature of Prinz's notion of *concern* is that it is a third-person response; whereas he takes *empathy* to be first-person response. Here again, I disagree with Prinz. The notion of care (concern in Prinz' terms) that I am employing is compatible with it being both a third-person response (abstract care) and a first-person response (we can call this empathic care) at the same time and separately at different times. It is compatible with my account that valuing the well-being of a target be the result of an agent having beliefs about what a target *should* be experiencing. An agent need not attempt to empathize with a target in order to care for that target. In such an instance, an agent would care for a target from a third-person point of view perhaps as a result of the ethical principles or duties. Crucially, however, on my account, care can also be a first-person response. To see how this is the case, let us re-examine the example of the agent interacting with a drug addict who is using a drug.

On my account, when an agent empathizes with the drug addict, the agent will experience an emotional state that is congruently valenced with that of the target. So the agent may experience pleasure when the drug addict is experiencing relief upon using a drug. Even though the agent arguably *should not* feel pleasure, what is important is that both the agent and the target experience a positively valenced emotional state. In addition, the agent can take the target's perspective and become aware of the drug addict's broader concerns such as: sharing the pleasure of drug use with another person, quitting drugs altogether, finding support in quitting and so on. Finally, when accurately empathizing, the agent must also feel motivated to help the target with these concerns. That an agent cares for the target is what explains the agent's motivation to help the target with these

[23] This example of the winning runner is one that I will return to below.

concerns at the particular times when the agent is accurately empathizing with the target. Care, in this instance, is a first-person response. Even though the agent did not care for the well-being of a target *prior to* perspective-taking, an agent can attempt to empathize with a target and, when successful, feel motivated to help that target with their concerns. On my account, care for the drug addict does not (always) result from an understanding of what the drug addict *should be* experiencing. Rather, it can be the result of the agent assigning value to the well-being of the drug addict based on the agent's changing awareness of the drug addict's changing concerns. Care can be first person-response resulting from empathy.

To repeat, on Prinz's matching account, care for a target is the result of abstract third-person reasoning about what a target should feel. Upon seeing a drug addict use a drug, an agent may share the emotional experience of the target. In this case, both the agent and the target will experience pleasure. But for Prinz, this is where empathy and care part ways. When the agent begins to care for the target, the agent's emotional state changes from pleasure to worry (or other emotions of disapprobation such as anger or disgust). The agent becoming aware of what the target should be experiencing causes this change. And it is separate from what the agent is experiencing and from what the target's concerns are. Care, for Prinz, does not result from taking a target's perspective. It is not a first-person response. Hence, he concludes that it should be treated separately from empathy.

I agree with Prinz insofar as it is possible to care for (or as Prinz puts it be concerned for) a target without empathizing with that target. But I also think (as I have argued above) that care influences empathy both prior to and subsequent to the activation of empathy's component processes. And in the case when care results from perspective-taking, care explains why agents who attempt to empathize with a target feel motivated to help that target more often than agents who do not attempt to empathize with the same target. Thus, care and empathy are not always separate phenomena. Accordingly, they should *not* always be treated separately.

I have argued that care and concern are the same insofar as they both involve assigning value to the well-being of a target. But I have distinguished care from concern on the basis of two differences. First, care need not be always negatively valenced. And second, care need not always be a third-person response. We can conclude from this that assigning value to the well-being of a target occurs in empathy, and that for this reason, empathy and care for targets should not be analyzed separately. And as mentioned, even on a matching account, the match resulting from empathy often leads agents to feel motivated to help targets. An account of empathy that treats care separately does not explain this motivation. Care is an important component of empathy in just the same way that emotional contagion or perspective-taking is. Care shapes the motivational end state of empathy. And by treating care and empathy together we can better predict and explain an empathizing agent's behaviour.

Prinz's advocacy for the separation of care and empathy is presented in the context of his rejection of empathy being important for moral judgment, moral development, and moral motivation (Prinz 2011a, 2011b). As we have seen, for Prinz, when an agent empathizes with a target, that agent matches the content of the target's emotional states. But he argues that this phenomenon is *not* important for the development of moral behaviour or moral theories. Accordingly, he argues that we should not cultivate what he calls "empathy-based morality". He believes that negative emotions embodying judgments of disapprobation such as concern (expressed as worry), anger, disgust, shame, and so on, are more important for morality than empathy. Empathy is not required because it is a weak moral motivator, and because it is biased towards those targets that agents take to be similar to themselves.

In this context, we can better understand his prescription that empathy should be treated separately from care. However, to accept Prinz' conclusion that empathy is not important for morality, we would have to accept that empathy and care are not only methodologically separable, but that they also only occur separately. I have so far provided reason to believe that care and empathy should be treated together by describing the roles of care in shaping the component processes of empathy. For example, care affects when empathy occurs. And when an agent cares for a target as a result of

perspective-taking, it motivates helping behaviour. I will now argue that the roles of care in relation to empathy lend credence to the stronger claim that care is *constitutive* of empathy. If this claim is correct, then *empathy* will indeed be important for morality. Like Prinz, I think that *care* is important for morality. So rather than directly addressing Prinz's arguments "against" the importance of empathy for morality, I will show that insofar as the processes of empathy are both caused by and are often the cause of care, empathy as a whole will be important for morality. To show this, I will re-examine some of the component processes of empathy and present two criteria for evaluating whether care is constitutive of these processes. Although the component processes of empathy can occur independently from care, it will turn out that there good reason to claim that care is constitutive of empathy.

2.10 On the constitutivity of care in empathy

Let us briefly return to one of the reference-fixing examples of empathy that I presented at the outset of this chapter. Imagine that your friend tells you that they have just been fired from a job they found fulfilling. Upon hearing this, you feel disappointed along with your friend. You know a little about your friend's difficult relationship with their employer and this leads you to believe that they were dismissed unjustly. You also feel motivated to help your friend find a new job. In this example, where does empathy stop and care begin? Or vice versa, where does care stop and empathy begin? Some accounts of empathy accept that the boundaries between care and empathy are easily distinguishable and that empathy is merely a matter of an agent matching the relevant beliefs and or emotions of a target (Ickes 1993; Darwall 1998; Prinz 2011a, 2011b). But I am proposing an account on which care plays an important role in enabling empathy to start, and in regulating the motivations that agents have to help targets. By examining these roles and their relation to empathy's component processes, I will argue that care is constitutive for empathy.

Care influences when empathy occurs and influences the motivations to help that an agent experiences in empathy. If care is removed, then an agent's competence for

empathic accuracy is greatly diminished. Accordingly, the following two conditions provide a partial specification of the *constitutivity* of care in empathy.

1) That care is reliably and *typically* part of empathy.
2) That accurate empathy cannot occur *without* care also occurring.[24]

The questions then are 1) for which of the component processes of empathy is it the case that care is typically a cause or a result of that process, and 2) is it the case that when accurate empathy occurs, care co-occurs? In the next section, I examine how the component processes of empathy fare with regard to meeting each of conditions just listed.

2.10.1 Mindreading

Mindreading is the first component process of empathy that I will examine with regard to 1) whether care is typical of its occurrence, and 2) whether care necessarily co-occurs with it. Nichols and Stich (2003) describe mindreading as "the capacity for ordinary people to understand the mind". Similarly, Baron-Cohen (1995) calls mindreading an ability "to infer" what is on another's mind. The term is generally used to refer to our ability to attribute intentional states (e.g. beliefs and desires) to others about their behavioural goals. In the well-known "Sally-Anne" task, Baron-Cohen et al. measure this ability in children by specifically examining whether children can attribute false beliefs to targets in relation to the goal of finding a ball which has been moved from one basket to another (Baron-Cohen et al. 1985). The experimental procedure is as follows. Two dolls (Sally and Anne) are presented to a participating child. Sally places a marble in a basket in front of her and leaves the scene. Anne then takes the marble out of Sally's box and places it in her own. Sally returns, and the participating child is asked "Where will Sally look for her marble?" Children who fail this task by pointing to where the marble has been moved to (as opposed to where it had been before the move), are said to have difficulty mindreading (*Ibid.*).

[24] The first of these conditions is weaker than the second. It specifies that when empathy occurs, care will likely also occur.

A major debate in the literature on mindreading is about the *mental format* of the processes involved–whether they are *theory* based (Baron-Cohen et al. 1985; Carruthers and Smith 1996, Gopnik and Wellman 1995) or *simulation* based (Goldman 1992a; Heal 1996; Goldman 2006). In the former case, mindreading would be a matter of having a theory resulting in thoughts about a target's mental state, while in the latter it would be a matter of more emotionally and motivationally laden imaginings of what their mental state is like.

As earlier mentioned, this debate has been extended to encompass the question of whether the nature of empathic processes are theory based or simulation based (Currie and Jureidini 1995; Goldman 1992b; Gallese and Goldman 1998; Sorensen 1998; Adams 2001). Important in this debate is the claim that empathy is equivalent to mindreading, or that mindreading is a component process of empathy. For the purposes of this chapter, the debate about the format of the processes of mindreading can be set aside. The question presently at hand–of whether mindreading typically involves care–can be posed independently of both whether we take mindreading to count as empathy or whether we take it to be a component of empathy, and whether it is theory-based or simulation-based. We need only acknowledge that many theorists take mindreading to be an important component of empathy.

The question then is: does mindreading typically involve care? To answer to this question, let us look again to the Sally-Anne task. The task measures whether a participant can, in attributing a false belief to a target, understand the behavioural intent of that target. Care–as the value that an agent places on the well-being of a target–does not seem to be typical in this mindreading task. The task does not typically involve an inducement or an assessment of the participant's care for the dolls. This is especially true on an account of mindreading that claims mindreading is theory driven. On a "*theory-theory*" account of mindreading, what is important is that the agent have the right sort cognitive states representing a target's mental states. On the other hand, *simulationist* accounts of mindreading allow for a richer characterization. They include the emotions and motivations that an agent may have towards a target. Accordingly, it is more plausible to posit that care is typically involved in mindreading on a simulationist

account. But again, simulationist accounts do not cite evidence that could support the claim that care is typically or necessarily involved in mindreading.

Even though there is no evidence to suggest that care is typically involved in mindreading, we can still ask whether care is required for mindreading to occur. Due to fact that care does not seem to be *typically* involved in mindreading, I think we should also conclude that care is not *necessarily* involved. An agent does not need to value the well-being of a target in order to have an expectation about that target's behavioural intentions. A child participating in the Sally-Anne task may not care at all about either doll, but still be fully possessed of the ability to *mindread* insofar as that child is able to tell where the target doll will look for the ball after it has been moved to the second basket. Thus, care is not typically involved in mindreading, nor it is necessarily involved.

2.10.2 Emotional contagion

Is care typical of or necessary for *emotional contagion*? Emotional contagion is a phenomenon in which an agent involuntarily "catches" the emotional state of a target. If an agent witnesses a target falling off their bicycle, that agent may involuntarily feel a somewhat painful discomfort. The agent need not already care for the target for emotional contagion to occur. Emotional contagion can occur when an agent is not acquainted with the target, or when an agent is acquainted with and dislikes the target. However, as discussed, emotional contagion can lead to an agent gaining some information about the target's perspective. When the agent witnesses the target falling off a bicycle, the agent's experience of discomfort is an emotional state that may lead the agent to imagine what it would be like for the agent to also fall off a bicycle. That is, emotional contagion may lead to *like-me perspective-taking*. In this example, the agent's discomfort alone (resulting from emotional contagion) may not be sufficient to lead the agent to care for the target. But if this discomfort leads the agent to imagine *themselves* in such an uncomfortable situation, then the agent is likely to come to care for that target. Insofar as emotional contagion typically results in an awareness of (some of) a target's concerns and a motivation to help that target, we can say that emotional contagion typically involves care. But is it in fact typical that emotional contagion leads to such

awareness and motivation? This is an empirical question to which I find no evidence to support an answer. However, we can easily imagine cases where the states produced by emotional contagion are purposefully suppressed or are simply not deemed important enough at the time to attend to. In such cases, emotional contagion will not lead to perspective taking and in turn care. Thus, emotional contagion can occur without care.

2.10.3 Mimicry (behavioural and neural)

Mimicry often leads to shared emotional states. In this section, I will examine whether agents that mimic targets typically or necessarily place value on the well-being of those targets. Two types of mimicry often feature as component processes of empathy: *behavioural* and *neural*. I will address them in turn. *Behavioural mimicry* is a phenomenon whereby an agent involuntarily imitates the behaviour or adopts the posture of a target (Field et al. 1982, 1985; Meltzoff and Moore 1977; Stern 1977). In such cases, it is posited that the agent may come to experience a state that is similar to that of a target. I consider behavioural mimicry to be a sub-type of emotional contagion in that the state experienced by the agent is one which is involuntarily "caught". Accordingly, we may conclude, as in the case of emotional contagion, that behavioural mimicry typically involves care only in cases where such mimicry leads an agent to an awareness of a target's concerns and a motivation to help. If an agent attends to the states produced by behavioural mimicry, the agent may imagine what it would be like for that agent to be in the target's position. However, as a sub-type of emotional contagion, we may also conclude that care is not necessarily involved in behavioural mimicry. An agent can experience a state that is relevantly similar to that of a target as a result of behavioural mimicry, but that agent need not care for the target in order for this to follow.

The second type of mimicry is *neural mimicry*. A prominent account of empathy that appeals to neural mimicry is that of Preston and de Waal (2002).[25] Their "Perception-Action Model" is a *matching account* of empathy according to which empathy occurs when an agent neurally matches the "embodied representations" of a target (Preston and

[25] The activation of mirror neurons is also a form of neural mimicry. Accordingly, I take my conclusions about neural mimicry to apply to accounts that explain empathy in terms of mirror neurons.

de Waal 2002, p 42). This matching occurs involuntarily. And, importantly, the agent uses "information about the self… to model the states of others." (Preston and de Waal 2002, p 43). However, Preston and de Waal do not specify exactly what information about the self is used by the agent. This is crucial to assessing whether care is involved in neural mimicry on their account of empathy. Although an agent may involuntarily neurally mimic the emotional states of a target, care would only be involved in so far as the agent also mimics the value that the target places on their own well-being. If we assume that neural mimicry *does* involve the agent neurally mimicking the value that a target places on themselves, then we can say, by hypothesis, that neural mimicry typically involves care.[26] And it follows that an agent neurally mimicking a target's care for themselves cannot do so without also caring for that target (if only "offline" or in simulation). That is, if an agent *does not* mimic the value that an agent places on themselves, then only partial mimicry has occurred. Having stated this however, more empirical research is needed to better understand whether when an agent involuntarily neurally mimics a target, that agent also mimics the care that the target places on themselves. Thus, it remains inconclusive whether neurally mimicking a target typically involves care, or whether it can occur in care's absence.

2.10.4 Like-me perspective-taking

Stotland (1969) introduced the distinction between two types of perspective-taking which map on to what I have been calling "like-me" and "like-other" perspective-taking (Stotland 1969, p 288, 289). In Stotland's terminology they were called "imagine-self" and "imagine-him" perspective-taking (Stotland 1969). Like-me perspective-taking involves imagining how an agent would think and feel were they in a target's situation. It can be described as an agent "projecting" themselves, *as that agent*, into the situation of the other. It is important to emphasize that in *imagine-self* perspective-taking, the agent imagines what it would be like if that agent (as themselves) were in that target's situation. This is why I call this type *like-me perspective-taking*.

[26] It follows that if a target does not value their own well-being, this would also be mimicked. In such a case, care would not be involved.

In her experiments on like-me perspective-taking, Stotland was interested in the question of whether it had "empathetic emotional consequences as well as predictive value." (Stotland 1969, p 289). She found that when participants were instructed to engage in like-me perspective taking, as opposed to merely watching and focusing on a target's behaviour, there was an increased physiological and emotional response. Thus, we can ask whether such an emotional response shows that like-me perspective-taking typically involves care?

An examination of Batson's work will be of help in answering this question because it makes substantial use of Stotland's distinction between types of perspective-taking. I think that the work of Batson et al. shows that care is typically involved in like-me perspective-taking. Specifically, the care typically involved in like-me perspective-taking is an *egoistic* care for a target in that it is dependent on the agent first caring for themselves. Batson shows that when participants engage in like-me perspective-taking upon witnessing a target in distress, they feel upset, anxious, disturbed, and so on (Batson 1991, p 77). This, more often than not, motivates targets to relieve their own distress rather than that of the target (Batson 1991, p 78). But although agents that engage in like-me perspective-taking are less likely to help targets (as compared to agents engaging in like-other perspective-taking), their emotional responses resulting from like-me perspective-taking indicates that they care for the well-being of targets. They care for them in so far as they care for themselves. When an agent imagines themselves in a target's distressful situation for example, that agent feels distressed because that agent is aware that the situation the target is in is somehow harmful to their well-being. The agent cares about that target's well-being. However, this care that the agent has for the well-being of the target motivates the agent to relieve their own distress rather than that of the target. Thus, even though an agent may care for a target as a result of like-me perspective-taking, the motivation that this produces is not directed at the target. The motivation to help a target resulting from like-me perspective taking and care is self-directed at the agent, rather than other-directed at the target. Importantly, this provides evidence that care is nonetheless typically involved in like-me perspective-taking.

Let us now examine the second condition of care's constitutivity of empathy: whether like-me perspective-taking can occur without care. We have seen that when an agent engages in like-me perspective-taking, that agent cares for the well-being of the target insofar as that agent cares about themselves. If the agent imagines caring about themselves when engaging in like-me perspective-taking, the agent will care about the target. But it is possible for an agent not to care about themselves when engaging in like-me perspective-taking. Often agents engage in risky, reckless, or obviously harmful behaviour. In such cases, we can assume that agents have limited care or no care for themselves. Now if such an agent engages in like-me perspective-taking upon witnessing a target in distress, and imagines that they were in that target's situation, then that agent would probably not experience the feelings of personal distress similar to those reported in the experiments of Stotland and Batson. We can thus predict that if an agent that engages in like-me perspective-taking, and in doing so imagines themselves not caring about themselves, then that agent would also likely not care about the target. It follows that a requisite for care being involved in like-me perspective-taking is that the agent would care about themselves were they in a situation like that of a potential target of perspective-taking. This indicates that like-me perspective-taking can occur without care, and is therefore *not necessary* for like-me perspective-taking to occur.

2.10.5 Like-other perspective-taking

The second type of perspective-taking is what Stotland (1969) calls "imagine-him" or what Batson (1991) calls "imagine-other" perspective-taking. This second type characterizes the act of imagining what and how a target is thinking and feeling in a situation with a voluntary effort to concentrate on what it would be like for the target (not the agent themselves) to be in that situation. This type of perspective-taking takes into consideration beliefs that the agent has about a particular agent. Potential examples of such beliefs include beliefs about a target's character, their present context, their history, their values, and so on. It is important to note that what differentiates this type of perspective taking is that an agent imagines what it would be like to be a particular target, rather than a more general like-me imaginative act that would involve an agent making

use of their familiarity with a situation and their prototypical beliefs about what they would experience were they in a target's situation. This type of perspective-taking is richer in that it not only takes into consideration what it is like for me (as an agent) to be in a target's situation. It also takes into consideration information that agents have about particular targets.[27] This is why I call this type *like-other perspective-taking*.

Similar to like-me perspective-taking, like-other perspective-taking involves considering the effects of a target's context (or environment) on that target. But dissimilarly, like-other perspective-taking results in less personal distress than like-me perspective-taking (Stotland 1969, p 297). And it is well empirically established that agents engaging in like-other perspective-taking are much more likely to be motivated to help targets of empathy (Batson 1991; Batson et al. 1997; Lamm et al. 2007). Does like-other perspective taking typically involve care?

Indeed, like-other perspective-taking is the phenomenon that presents the strongest evidence for care being typically involved. There is a high correlation between like-other perspective taking and helping behaviour (Batson 1991). And this correlation is explained by the role of care. It is the care that an agent has for a target that motivates an agent to engage in the time-consuming act of like-other perspective-taking. The aforementioned type of *like-me perspective-taking* can be performed voluntarily. But like-me perspective-taking does not require an agent to have a rich set of beliefs about a target. The agent can recruit their own familiar experiences, and expectations to imagine the thoughts, feelings, and concerns of a target. Or an agent can use these to imagine what it might possibly be like for *that* agent to be in a prototypical conception of their situation. On the other hand, when an agent engages in *like-other perspective-taking*, that agent must either have beliefs about a target's particular life, and beliefs about how that particular target experiences their environment. In performing this imaginative act, an agent may rely on beliefs acquired by prior acquaintance with the particular target. But often, an agent will need to acquire these beliefs by lengthy observation and interaction. It is this voluntary and time consuming effort involved in like-other perspective taking that provides support

[27] And this requires agents to have longer periods of interaction with potential targets of perspective-taking in order to imagine what it is like to be in a target's *particular* situation.

to my claim that it *typically* involves care. At times, an agent may be motivated by other values when engaging in like-other perspective taking. But I think it is correct to say that in most cases, the value of care plays an equally or more important motivational role.

Turning now to the question of whether like-other perspective-taking can occur without care, we can say that it can get started, but without care empathic accuracy would be greatly diminished or be impossible. Caring for a target while engaging in like-other perspective-taking enables the agent to become aware of differences between how that agent would care for themselves in a particular context as compared to how the target's value of care may be conceived of differently in the same context.[28] It may also lead an agent to become aware of differences between how that agent is experiencing their context in relation to this different conception of care. Furthermore, care contributes to motivating an agent to focus on their awareness of a target's concerns, rather than on their own personal distress. This, in turn, explains why agents engaging in like-other perspective-taking are more likely to help targets of empathy.

2.10.6 Summarizing constitutivity

The following table will summarize the conclusions presented above:

Table 1: *Component processes of empathy compared to typicality and necessity of care*

	Mindreading	**Emotional contagion**	**Behavioural mimicry**	**Neural mimicry**	**Perspective-taking (like-me)**	**Perspective-taking (like-other)**
Care typical	No	Yes	Yes		Yes	Yes
Care necessary	No	No	No		No	No
Evidence of constitutivity	No	Some	Some	Inconclusive	Some	Yes

[28] I will expand on the consequences of different conceptions of care in the next sections.

Having identified which of empathy's component processes meet the criteria of (1) *typically* involving care, and (2) not occurring without care (or *necessarily* involve care), we can draw the conclusion that *care is constitutive of empathy*. Applying the first criterion to the component process of empathy yielded the result that *care* is highly typical of empathy. And applying the second criterion showed that without care, the occurrence of empathy would be less frequent, and empathic accuracy would be diminished to the point of being impossible while engaging in like-other perspective-taking. Care plays important roles in almost all of empathy's component processes. And without it, the activation and accuracy of these processes would be greatly diminished. To treat empathy and care separately, as Prinz does, is to unjustifiably separate empathy from component processes that are both typical and necessary to its end state when it occurs. Such an approach offers no explanatory benefit, and may lead to misguided empirical investigation. Accordingly, I take care to be *constitutive* of empathy. As I have just argued, this does not mean that care is necessary for the component processes to get started. Care is a necessary motivational state of an agent when the component processes of empathy result in accurate empathy. Batson also argues for a similar conclusion (Batson 2011). He states that for empathy to occur, more is required than an agent perceiving a target as being in need and becoming aware of the target's needs (i.e. concerns). In addition, the agent must care for the target as a condition for feeling motivated to help the target with their concerns (Batson 2011, p 41).

Taking the value of care to be constitutive of empathy opens a space to ask important questions about the roles of values (other than care) that agents and targets have. As an individual value, care for the well-being of a potential target is importantly connected to other values that both agents and targets have, and to broader social values that may be expected or imposed by the context in which agents and targets engage in empathy. In the next section, I will address how empathy is shaped by the relationship between care, other individual values held by agents and targets, and by broader communal and institutional values that are at work in the contexts in which empathy occurs.

2.11 The roles of "values" in empathy

So far, I have argued that the value of care is constitutive of empathy. In doing so, I stated that the value an agent assigns to the well-being of a target affects whether empathy will occur or not. But in addition to being influenced by care, empathy is influenced by other values. I will begin this section by providing some general remarks about *what* values are in relationship to *care* and *motivation* in empathy. Then I will describe three ways that values influence empathy. First, values influence when empathy "gets started"; they influence what counts as a *triggering cause* of empathy. Second, values generally (like care) make empathy what it is; they *structure* the component processes of empathy. And third, values affect empathy as a motive; they influence emotions and behavioural motivations in empathy. Describing the influence of values generally on empathy (constituted by care) will allow me to provide a sketch of *how empathy works* on a care account. Doing so will involve sketching a model of empathy by reference to the interactions between *values*, *motivational orientations*, *environmental contexts*, and *goals*.

2.12 Care and other values

Accounts of empathy have largely ignored the role of values generally. One reason for this is conceptual. Theorists in many disciplines speak about the importance of values. But often, they use the term while saying very little about what they take values to be. Interpreting what values are is often left to the reader's familiarity. Another reason why values are not addressed in accounts of empathy is that values are difficult to operationalize (Gecas, 2008). They are often set aside in favour of more easily measured variables such as "attitudes", "norms", and "roles" (*Ibid.*). Furthermore, there is a tendency in science and philosophy of science to downplay the role of values as outside the proper domain of scientific investigation (Graham 1981; McMullin 1982).[29] Many researchers have taken science to provide objective and universal explanations. However,

[29] More recently, many philosophers of science have acknowledged and described the roles of values in science. They argue that values play important roles in influencing the goals of science, theory choice, and hypothesis acceptance. For a review, see Graham (1981) and McMullin (1982).

my view is that integrating the role of values in empathy is important for providing a fuller and contextually-sensitive account.

As a value in empathy, care is a *terminal value*–a belief that an agent has about a "desirable end-state of existence" for a target in a particular context (Lovejoy 1950, p 596-597; Rokeach 1973, p 7).[30]

As a *terminal value*, care interacts in important ways with *instrumental values*. For example, let us take empathy as a motive to helping behaviour. When an agent empathizes with a target, that agent's behavioural motivations will be influenced by *instrumental values*–beliefs about contextually appropriate "modes of conduct" (Lovejoy 1950, p 596-597; Rokeach 1973, p 7). Having taken the perspective of a target who is asking for money on the street, an agent may be motivated to help that target by giving them money. This agent's actions can be interpreted as being influenced by the instrumental value of "*charity*". A different agent, having taken the perspective of the same target may feel motivated to help by ignoring the target and walking on. This latter agent's actions can be interpreted as being influenced by the value of "*personal autonomy*". For present purposes, I will leave the question of whether accurate empathy has occurred in either case. What is important to emphasize is that an agent's values will influence the behavioural motivations that an agent has when empathizing with a target.

Values, like desires, are motivational. It is important to note that, although values may cause desires, they are distinct from desires (Williams, 1968, p 284). For example, I may desire to eat given that I value food. But my desire to eat food is not equivalent to my valuing food. Valuing food is prior to desiring food. The value I have is FOOD, according to which I believe it would be desirable for me to be in a context where food is present (then I would eat it).

---

[30] I follow Rokeach in taking values to be *beliefs* that have a cognitive, affective, and behavioural component. But it is important to note that, unlike other beliefs, values are *evaluative* in that they are about states or behaviours that are *desirable* or *undesirable*, rather than *true* or *false*. It is also important to note that a terminal value is not a description of an end state. Accordingly, care in empathy does not determine an agent's conception of the context or conditions for a target's well-being. It specifies that in a particular end-state, a target's well-being is desirable. Though care is of central importance to empathy, it is but one value among an agent or target's total system of values (Williams and Albert 1968, p 287).

Values may be embedded in verbal behaviour, actions, and environmental contexts (Williams and Albert, p 287). In the case of values being embedded in environmental contexts, what I mean is that environmental contexts impose certain values on agents due to the conventions that are associated with desired outcomes or behaviours in those contexts, and due to the structural features of the physical environment. An agent need not explicitly endorse the values that are influencing or guiding their behaviour. An agent may verbally express the values they *conceive* to be desirable, but acknowledge that context requires that they *operate* according to different values (Morris 1956a, 1956b; McLaughlin 1965). Likewise, an agent may act according to values that they are not aware of or that they do not express (*Ibid.)*). Accordingly, it is not surprising that an agent's values are often inconsistent. Any agent (or target) may have values that are mutually inconsistent at a particular time, or which become inconsistent depending on context and development. For example, the value of "money" and the value of "convenience" are often inconsistently held by the same agent purchasing a meal at a fast-food restaurant. Such v*alue trade-offs* indicate that agents frequently have inconsistent values (Tetlock 2003).

It is also important to note that any particular value is not independent from other values. An agent's values influence each other. For example, an agent in a particular context may value "freedom", and that value may be influenced by that agent also valuing "adherence to social hierarchy" (Williams 1968, p 286). Thus, the agent's *terminal value* of "freedom" is shaped and constrained by the *instrumental value* of "adherence to social hierarchy". Another agent in a different context may also value "freedom". But in this agent's context, the agent's value "freedom" is influenced by their valuing "equality" (as opposed to "hierarchy") (*Ibid.)*. In either case, the agent's value of "freedom" will be influenced by other values, and thus shape the agent's experience differently (*Ibid.*).

Given these general remarks about value, we can begin to see that both terminal values (e.g. care) and instrumental values (e.g. appropriate behavioural motivations) will interact and influence each other in significant ways in empathy. I will argue that values affect when empathy occurs, what empathy is, and the behavioural motivations that agents experience when empathizing with a target. Before describing the three roles of values on

my care account of empathy, it will be useful to summarize some important aspects of my account and how they relate to empathic accuracy.

2.13 A care account and empathic accuracy revisited

For accurate empathy to occur on my care account of empathy there are four criteria:

1) Agent is *aware* of a target's *concerns*
2) *Matching emotional valences* between agent and target
3) Agent *cares* for the target
4) Agent feels motivated to help target with concerns

The first criterion requires that the agent become aware of a target's concerns. This awareness can occur via involuntary component processes such as emotional contagion or neural mimicry; or via component processes under more voluntary control such as perspective-taking. For example, upon hearing from a friend that they just lost their job, an agent could become aware of that target's concerns such as paying their debts, and finding satisfaction in a different line of work.

The second criterion is that there be a match between the valences of the emotional states of the agent and the target. This valence matching does require that both agent and target experience emotional states with the same content. For example, upon seeing that a friend just win a race, an agent could match the emotional state valences of the target (at a particular time). The runner may be experiencing joy, whereas the agent may be experiencing pride. What is important is that there be a *valence match*, as opposed to a *content match*.

The third criterion is that an agent cares for a target. I have characterized this care as a value that an agent assigns to the well-being of a target. For example, upon seeing a person sitting on the sidewalk asking for money, an agent could care for that target. The agent could conceive of an environment in which the target's concerns are met, and experience this conception as desirable. When accurate empathy occurs, an agent places value on the well-being of a target.

Fourth, for accurate empathy to occur, an agent must feel motivated to help a target with their concerns. It is important to emphasize that the agent merely needs to *feel* motivated to help the target with their concerns insofar as the agent has conceived of behaviours which will contribute to this goal. The agent need not act on these motivations.

In sum, matching accounts of empathy involve an exact match between the state of an agent and that of target. Concern accounts involve a partial match between agents and targets in terms of the valence of their emotional states. Concern accounts also add a motivational component whereby for empathy to occur its end state must include an agent that is motivated to help a target with their concerns. Further, the care account adds that an agent's care for the well-being of a target is what explains this motivation. In support of this claim, I have shown how care is typically involved in many of empathy's psychological processes, and how the value of care is a necessary component of empathy's end state. With these criteria in mind, let us move on to examining the three roles that values (generally) play in empathy.

2.13.1 Values influence what causes empathy (i.e. they are triggering causes)

Another agent in the environment or a "target", is commonly the *triggering cause* of empathy (Dretske 1991). I am following Dretske here in distinguishing between triggering and structuring causes (Dretske 1991, ch. 2). Targets are what trigger the activation of empathy's component processes. Which targets are taken by an agent as a potential target of empathy is the result of a confluence of factors. In some sense, the entirety of an agent's history, environment, and psychology can be understood as contributing to which targets an agent empathizes with. This is because an agent's psychological development generally will contribute to explaining this phenomenon of selection. Also, an agent's interactions with a particular target often vary over time, and will thus influence whether that particular target is taken as a potential target of empathy. In modeling these interactions I will examine several factors, which along with values, influence which targets become *triggers* of empathy. Values influence which targets will be taken as potential targets of empathy [PTE]. But their influence is *indirect*. There is no immediate causal connection between an agent's values and which targets an agent takes

to be PTE. I will argue that it is more accurate to describe what determines which targets are selected as PTE as a system of reciprocal causal relationships between an agent's *terminal values, goals, motivational orientations, environments,* and *instrumental values*. In this section I will address the first four of these factors. In the next section on values as structuring causes of empathy, the role *instrumental values* will be described.

As discussed above, *terminal values* are an agent's beliefs about a desirable end-state of existence. Examples of terminal values are "freedom", "equality", and "a comfortable life" (Rokeach 1973, p 252; Curhan et al. 2006). Insofar as an agent values "freedom" that agent will assign value to an outcome in which freedom is present and promoted. That agent will feel positive emotions when freedom is promoted and negative ones when it is challenged. And that agent will strive for or behave in such as a way as to achieve an end state which that agent describes as "free". This brings us to the second factor: goals.

Agents engage in goal-directed behaviour (Adams and Enç 1992; Barker 2008). And terminal values influence these *goals*. Usually, behaviours will not be directed at attaining a particular valued end-state. Rather, an agent's goals are *influenced* by terminal values in that the latter affect which goals an agent chooses to pursue. For example, an agent who values "a comfortable life" will not go about aiming to achieve a comfortable life in all contexts. At any moment, the goal of agent's behaviours will not usually be to realize a terminal value. Rather, terminal values influence an agent's *choice* of goals. That is, the goals an agent chooses to pursue will be influenced by their belief that a comfortable life is desirable.

We are now beginning to approach the indirect influence of values on an agent's selection of PETs. Terminal values influence an agent's choice of goals. These goals, in turn, will affect an agent's *motivational orientations*. When interacting with targets, agents have different social motives or *orientations*. In game theory, these orientations play an important role in determining the way in which agents interact with targets, and the strategies they use in order to achieve goal-directed outcomes (McClintock 1972; Kuhlman and Marshello 1975; Kuhlman et al. 1986; McClintock and Liebrand 1988). Four motivational orientations are distinguished:

1) Individualism (maximization of agent's own gain)

2) Competition (maximization of difference between agent and target's gain)

3) Cooperation (maximization of joint gain)

4) Altruism (maximization of other's gain)

The goals that an agent pursues will shape the motivational orientations that an agent is in when interacting with PETs. And these goals will, of course, vary according to environmental context. For example, an agent who values "profit" may be working at a very profitable company. While at work, that agent will pursue a variety of work-related goals such as: impressing a client at a meeting with competitors; discussing the budget at a staff meeting; visiting a construction site. A *change in goal* can affect the *motivational orientation* that an agent is in when interacting with targets that the agent encounters. For example, when pursuing the goal of impressing a client, the agent may be in a *competitive* motivational orientation towards other salespeople. When pursuing the goal of negotiating a budget at the staff meeting, the agent may be in a *cooperative* motivational orientation towards co-workers. Accordingly, changes in *goals* can involve changes in the *motivational orientations*, which in turn determine the behavioural strategies that an agent employs when interacting with targets (*Ibid.*). The significance of motivational orientations for which targets are taken as PETs is that changes in motivational orientations will involve changes in which targets an agent *cares* for.

On this model, I predict that motivational orientations will be importantly related to *the value of care* that agents assign to targets. It is more likely that an agent will care for a target when in the *cooperative* or *altruistic* orientations (as opposed to the *individualistic* or *competitive* orientations). Because an agent is interested in maximizing *joint gain* when in a cooperative orientation this indicates that the agent may value the well-being of the target insofar as that target's well-being is compatible with that of the agent's well-being. The strategy of "joint gain" that is at work in the cooperative orientation implies that the agent, at the very least, is *concerned* (in Prinz and Darwall's third-person sense) for the well-being of the target. Furthermore, I think that in the *cooperative orientation* it

is likely that an agent's concern may cause or be accompanied by attempts at *like-me perspective-taking* resulting in a *first-person* awareness of a *target's concerns* and a motivation to help that target with their concerns. As we have seen, valuing the well-being of a target is typical of like-me perspective taking. In the *altruistic motivational* orientation, that an agent care for the target is all but guaranteed. In an altruistic orientation, an agent is both more likely to care for a target and more likely to attempt to empathize with that target. This prediction is supported by evidence of the high correlation between empathy and altruistic behaviour. I have argued that what explains this altruistic behaviour as a result of empathy is that care is constitutive of empathy. Thus, I predict that in the *altruistic motivational orientation,* agents will likely take targets as PETs and engage in like-other perspective-taking. The role of care in like-other perspective-taking explains the selection of PETs in an altruistic motivational orientation.

Figure 1: *Motivational orientations and terminal values*

I have been arguing that values (*terminal values* and "*care*") will play a crucial role in determining which targets are taken by agents as potential targets of empathy. An agent will be in a particular context as a result of terminal values. In this context, that agent will pursue goals which will be met by adopting motivational orientations of interaction towards targets. I have predicted that in motivational orientations of cooperation and altruism, an agent is more likely to assign value to the well-being of a target. This is because care plays important roles in causing an agent to empathize with a target. And with greater reason, because care is constitutive of empathy, an orientation in which an agent is likely to care for a target is one in which an agent is likely to empathize with that target.

2.13.2 Values influence why empathy occurs (i.e. they are structuring causes)

Values are *structuring causes* of empathy (Dretske 1991). We have seen that values indirectly influence which targets (as triggering causes) an agent takes to be potential targets of empathy. On the other hand, a structuring cause is a cause which is responsible for a process *being* a process with a certain *product* (*Ibid.*). A product describes an event or condition, such that, until it occurs, the process has not ended (Dretske 1991, p 44). The product of empathy then–the point at which the process of empathy ends–is an agent's motivation to help a target with their concerns. Accordingly, the value of care and terminal values are *structuring causes* of empathy. They are the *causal conditions* for empathy being a process that causes an agent to feel motivated to help a target. We can schematize this relationship as follows:

Target (T) *causes* Empathy (E) which *causes* Helping behaviour (H)

We have already seen that terminal values and care influence which targets (T) cause empathy (E) to start. If we have knowledge of an agent's terminal values, we may have knowledge of that agent's goals, and the motivational orientations that the agent is in when pursuing those goals. This would allow us to infer from observation that a target (T) caused an agent to empathize (E) with that target, which in turn caused that agent to help (H) that target with their concerns. The agent helped the target when, and because, it saw a target which it selected as a potential target of empathy. But why did Empathy (E) cause helping behaviour (H)? This question is about the structuring cause (as opposed to the triggering cause) of the relationship between (E) and (H). It is about the conditions in which (E) causes (H) rather than something else.

Here again *care* will be important. Care affects which targets act as triggers of empathy. But it is also responsible for empathy being a process that causes helping behaviour–it is a structuring cause. When an agent is triggered to empathize by a target, *care* explains why a target feels motivated to help a target with their concerns as opposed to say feeling motivated to ignore the target or feeling motivated to perform some other activity. The

value of care that an agent assigns to a target explains *why* an agent feels motivated to help a target. This explanation of helping behaviour as a result of empathy resides not in the stimulus (or triggering causes), but in the correlations between care and helping behaviour. Care in empathy causes helping behaviour:

> Target (T) *causes* Empathy (E) which *causes* **Care (C) as a condition** for helping behaviour (H)

Targets (T) as triggering causes explain why empathy gets started *now*.[31] But care (C) as a structuring cause explains why helping behaviour results from empathy as opposed to something else.

Similarly, *instrumental values* are structuring causes of empathy. Instrumental values are desirable properties that an agent assigns to a behaviour. Examples of instrumental values include "respect", "self-control", "forgiveness", "ambition", "responsibility" (Rokeach 1973; Curhan 2007). Accordingly, like care, instrumental values are responsible for empathy being a process that results in a motivation to help a target. Instrumental values contribute to the explanation of why an agent who empathizes with a target helps that target with their concerns. It is because instrumental values are desirable properties of behaviour that they are motivational. For example, an agent that values "self-control" more than "forgiveness" may not help a target even if that agent cares for the target. On the other hand, an agent that values "responsibility" more than "self-control" is more likely to help a target (especially if the agent also cares for the target). Instrumental values do not reside in the stimulus or triggering causes of empathy. Like the value of care, they structure the process of empathy by being conditions for when helping behaviour will occur (as opposed to some other behaviour):

> Target (T) *causes* Empathy (E) which *causes* **Care (C) and Instrumental value (I) as conditions** for Helping behaviour (H)

[31] We have seen that care has an indirect influence on which targets are taken as triggers. But caring for a target is not itself the stimulus that causes empathy. Care is not itself a triggering cause.

Instrumental values are causal conditions or structuring causes of empathy that importantly contribute to explaining *why* empathy is a process that causes helping behaviour. They are distinct from triggering causes which explain *what* causes empathy to start at a particular time, in a particular context, when stimulated by a particular target.

2.13.3 Values influence how empathy operates (i.e. which motives are involved)

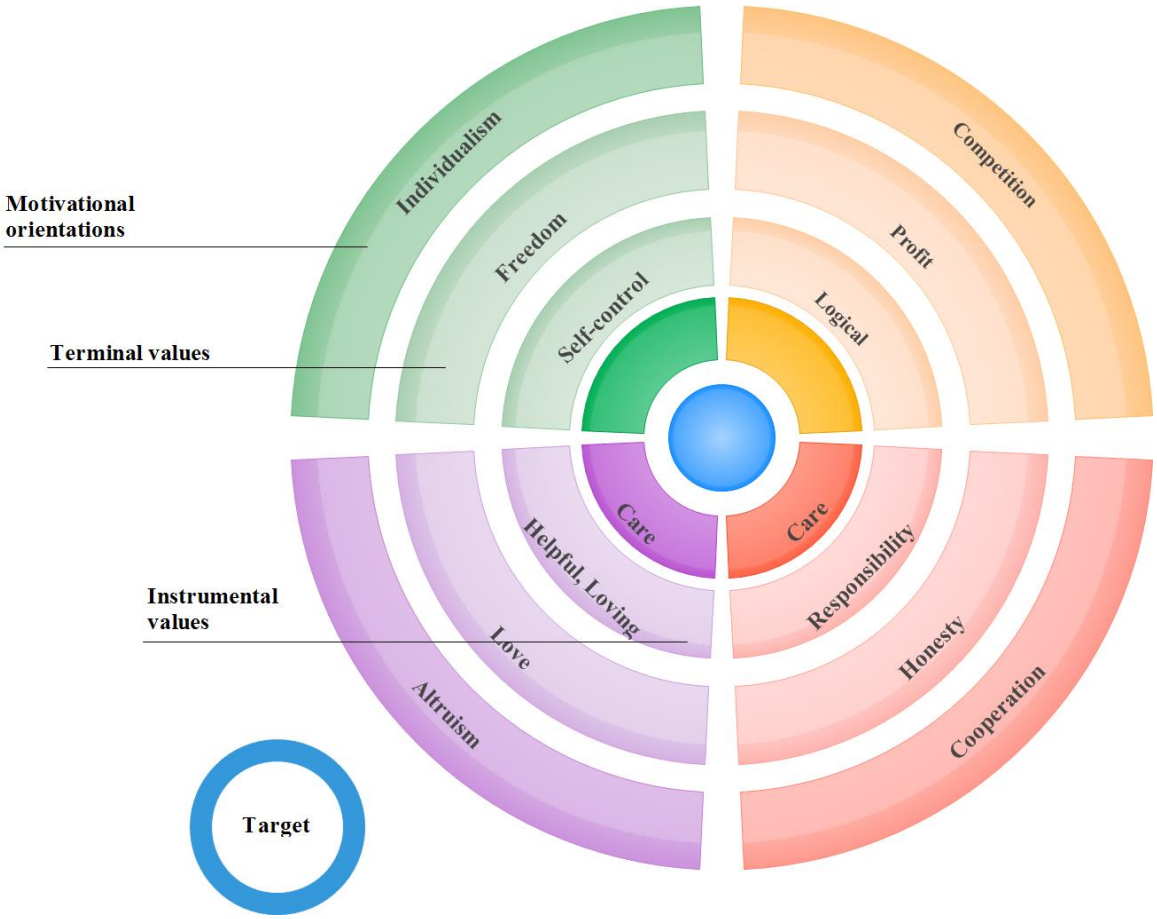
We have seen that instrumental values shape what agents take to be "desirable modes of conduct" (Rokeach 1973, p 7). They are properties of the *means* by which agents strive for desirable *ends*. Their significance for empathy is that they influence *how* empathy works. Specifically, instrumental values influence which behaviours an agent attempting to empathize with a target feels motivated to perform. How they do so is by their connection to *motivational orientations*.

I described above the role of motivational orientations in the selection of which targets an agent takes to be potential targets of empathy. These motivational orientations play another important role in empathy. Namely, they determine which instrumental values an agent will make use of in any given situation. This is because the motivational orientation that an agent is in subsumes different instrumental values. For example, when in the competitive motivational orientation, an agent's behaviours may be shaped by instrumental values such as "ambition" or "logical" (Rokeach 1973, p 119). When an agent is in the cooperative or altruistic orientation, an agent's behaviours may be shaped by values such as "broadminded" or "forgiving" (*Ibid.*). Thus, the motivational orientation that an agent is in will affect the instrumental values that shape an agent's behaviour. The import for empathy of the relationship between *motivational orientations* and *instrumental values* is that an agent's *motivations to help* a target with their concerns will depend both on the motivational orientation the agent is in and on the terminal values of that agent subsumed under that orientation.

Instrumental values are important to integrate into an account of empathy for their explanatory contribution. By examining the relationship between an agent's motivational

orientations and that agent's terminal values we can understand why an agent's behavioural motivations to help a target with their concerns are as they are, as opposed to being some other way. For example, $agent_1$ is in an altruistic motivational orientation when attempting to empathize with $target_1$ who is sitting in the street asking for money. In this orientation, $agent_1$'s behaviours are influenced by instrumental values such as "helpful" and "loving". This will contribute to an explanation of why $agent_1$ stops, talks to $target_1$ about their concerns and gives $target_1$ money. An agent with different instrumental values who is attempting to empathize with the same target may behave quite differently because this agent's motivational orientation is associated with different instrumental values. This $agent_2$ is also in an altruistic orientation and attempting to empathize with $target_1$. But $agent_2$'s altruistic motivational orientation is associated with instrumental values such as "independent" and "responsible". Thus, $agent_2$ may stop and talk to $target_1$, but decide not to give $target_1$ money. It is the difference between $agent_1$ and $agent_2$'s instrumental values under the same motivational orientation that explains the difference in the behaviours they each felt motivated to perform when attempting to empathize with $target_1$.

Figure 2: *Motivational orientations and instrumental values*



But this is not the whole story. We can imagine a case where $agent_1$ and $agent_2$ are both in the same motivational orientation (e.g. altruistic), and that this orientation subsumes the same instrumental values (such as "helpful" and "loving") for both agents. But even by keeping the relationship between motivational orientations and instrumental values constant, we can still imagine that $agent_1$ and $agent_2$ will differ in the behaviours they feel motivated to perform in order to help $target_1$. This is because *terminal values* also affect the behavioural motivations of agents. They do so *indirectly* by affecting an agent's conceptions of a target's concerns. For example, when engaging in like-me perspective-taking an agent will become aware of and attribute concerns to a target. And we have seen that this will involve a match in emotional valence between agent and target based on an agent's familiarity with contexts similar to that in which the target finds themselves. Importantly, the *content* of the concerns that an agent takes a target to have

will be influenced by that agent's terminal values. An agent that is taking the perspective of a target and becoming aware of that target's concerns sees the target in a particular context, makes use of past experience in similar contexts, and matches the valences of that target's emotional states. But this process is not *psychologically isolated* from terminal values. I think it is useful to think of the contribution of terminal values as an agent's "ideals" or "ideal outcomes" to the construction of an agent's awareness of a target's concerns. When attributing concerns to a target, an agent will be influenced by what that agent takes to be positive or ideal outcomes for that agent. As we have seen, this is because, when empathizing with a target, an agent cares for that target's well-being. Accordingly, we can say that how an agent conceives of a target's well-being is influenced by an agent's terminal values. Agents will differ in their terminal values, and so they will differ in how they conceive of the well-being of a target. In turn, different agents will construct different concerns that they attribute to PETs.

Returning to our example then, we can imagine two agents that share the same motivational orientation towards a target and whose motivational orientations subsume the same terminal values. And we can imagine that each of these agents will have different behavioural motivations to help the target with their concerns. This can be explained by the different concerns that each agent has attributed to the target. $Agent_1$'s conception of the well-being of $target_1$ will have been influenced by different terminal values than $agent_2$'s. All things being equal, $agent_1$ and $agent_2$ will have different behavioural motivations to help $target_1$ when attempting to empathize. And this will have been a consequence of the influence of terminal values on each agent's construction and attribution of concerns to the target.

We can now see why the care account of empathy that integrates the role of care and values more generally is an improvement over accounts of the matching account type. There is clearly a necessary evaluative component that influences when an agent will, as a result of empathy, feel motivated to help a target. And when this evaluative component is understood as preferences over possible outcomes for the target, we can see that these preferences can only be known by adopting the evaluative stance of the target. By doing so, the agent is motivated to act in the target's interests. This is what explains the

correlation between matching a target's internal state and helping behaviour. Without considering the role of care and values, matching accounts cannot explain this correlation.

2.14 Concluding remarks

This paper has been concerned with classifying what empathy is, and with proposing an account of how empathy works that better explains its motivational component. I started by classifying accounts according to their conception of empathic accuracy. Namely, by distinguishing between *matching accounts* and *concern accounts*. I then argued that concern accounts better explain how empathy works by including the component in empathy that motivates and agent to help a target with their concerns. Further, I built upon the work of concern accounts to present the *care account* of empathy. I argued that the care account is an improvement on existing concern accounts because it better explains and predicts empathy. The care account includes the roles of care and values more generally. It is a first step towards providing more detail about the causes of empathy, the conditions in which it occurs, and the motivations of agents and targets that cause helping behaviour. If the care account is correct in its central claims that *values*, *environmental contexts*, and *motivational orientations* are important to empathy, then going forward there is reason to reject matching accounts of empathy, and reason to favour concern accounts that integrate the factors that the care account emphasizes.

References

Adams, Frederick. “Empathy, Neural Imaging and the Theory versus Simulation Debate.” *Mind & Language* 16.4 (2001): 368–392. Print.

Adelmann, Pamela K., and Robert B. Zajonc. “Facial Efference and the Experience of Emotion.” *Annual review of Psychology* 40.1 (1989): 249–280. Print.

Anderson, Steven W. et al. “Impairment of Social and Moral Behavior Related to Early Damage in Human Prefrontal Cortex.” *Nature neuroscience* 2.11 (1999): 1032–1037. Print.

Barker, Gillian. “Biological Levers and Extended Adaptationism.” *Biology & Philosophy* 23.1 (2007): 1–25. *CrossRef*. Web.

Baron-Cohen, Simon. *Essential Difference: Male and Female Brains and the Truth about Autism*. Basic Books, 2003.

---. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT press, 1995.

Barrett-Lennard, Godfrey T. “The Empathy Cycle: Refinement of a Nuclear Concept.” *Journal of Counseling Psychology* 28.2 (1981): 91. Print.

Batson, C. Daniel. *Altruism in Humans*. Oxford; New York: Oxford University Press, 2011. Print.

---. “Prosocial Motivation: Is It Ever Truly Altruistic?” *Advances in experimental social psychology* 20 (1987): 65–122. Print.

---. *The Altruism Question*. Hillsdale, NJ: Erlbaum, 1991. Print.

---. “These Things Called Empathy: Eight Related but Distinct Phenomena.” *The Social Neuroscience of Empathy*. Ed. Jean Decety and William Ickes. Cambridge, MA, US: MIT Press, 2009.

Batson, C. Daniel, Shannon Early, and Giovanni Salvarani. “Perspective Taking: Imagining How Another Feels versus Imaging How You Would Feel.” *Personality and social psychology bulletin* 23.7 (1997): 751–758. Print.

Batson, C. Daniel, and Laura L. Shaw. “Evidence for Altruism: Toward a Pluralism of Prosocial Motives.” *Psychological Inquiry* 2.2 (1991): 107–122. Print.

Bavelas, Janet B. et al. “‘I Show How You Feel’: Motor Mimicry as a Communicative Act.” *Journal of Personality and Social Psychology* 50.2 (1986): 322. Print.

Blair, R. James R. “A Cognitive Developmental Approach to Morality: Investigating the Psychopath.” *Cognition* 57.1 (1995): 1–29. Print.

Blair, R.J.R. “The Amygdala and Ventromedial Prefrontal Cortex in Morality and Psychopathy.” *Trends in Cognitive Sciences* 11.9 (2007): 387–392. *CrossRef*. Web.

---. “The Roles of Orbital Frontal Cortex in the Modulation of Antisocial Behavior.” *Brain and Cognition* 55.1 (2004): 198–208. *CrossRef*. Web.

Borke, Helene. “Interpersonal Perception of Young Children: Egocentrism or Empathy?” *Developmental psychology* 5.2 (1971): 263. Print.

---. “Interpersonal Perception of Young Children: Egocentrism or Empathy?” *Developmental psychology* 5.2 (1971): 263. Print.

Carruthers, Peter, and Peter K. Smith. *Theories of Theories of Mind*. Cambridge University Press, 1996.

Chapman, Michael et al. “Empathy and Responsibility in the Motivation of Children’s Helping.” *Developmental Psychology* 23.1 (1987): 140. Print.

Charland, Louis C. “Emotion Experience and the Indeterminacy of Valence.” *Emotion and consciousness* (2005): 231–254. Print.

---. “The Heat of Emotion: Valence and the Demarcation Problem.” *Journal of consciousness studies* 12.8-10 (2005): 82–102. Print.

Colombetti, Giovanna. “Appraising Valence.” *Journal of Consciousness Studies* 12.8-10 (2005): 103–126. Print.

Curhan, Jared R., Hillary Anger Elfenbein, and Heng Xu. “What Do People Value When They Negotiate? Mapping the Domain of Subjective Value in Negotiation.” *Journal of Personality and Social Psychology* 91.3 (2006): 493–512. *CrossRef*. Web.

---. “What Do People Value When They Negotiate? Mapping the Domain of Subjective Value in Negotiation.” *Journal of Personality and Social Psychology* 91.3 (2006): 493–512. *CrossRef*. Web.

Currie, Gregory, and Jon Jureidini. “Delusion, Rationality, Empathy: Commentary on Martin Davies et Al.” *Philosophy, Psychiatry, & Psychology* 8.2 (2001): 159–162. Print.

Damasio, Antonio R. *Descartes’ Error: Emotion, Reason, and the Human Brain*. New York: Avon Books, 1995. Print.

Darwall, Stephen. “Empathy, Sympathy, Care.” *Philosophical Studies* 89.2 (1998): 261–282. Print.

Decety, J. “The Functional Architecture of Human Empathy.” *Behavioral and Cognitive Neuroscience Reviews* 3.2 (2004): 71–100. *CrossRef*. Web.

De Vignemont, Frederique, and Tania Singer. “The Empathic Brain: How, When and Why?” *Trends in Cognitive Sciences* 10.10 (2006): 435–441. *CrossRef*. Web.

Dimberg, U., M. Thunberg, and K. Elmehed. “Unconscious Facial Reactions to Emotional Facial Expressions.” *Psychological Science* 11.1 (2000): 86–89. *CrossRef*. Web.

Doherty, R. William. “The Emotional Contagion Scale: A Measure of Individual Differences.” *Journal of nonverbal Behavior* 21.2 (1997): 131–154. Print.

---. “The Emotional Contagion Scale: A Measure of Individual Differences.” *Journal of nonverbal Behavior* 21.2 (1997): 131–154. Print.

Dretske, Fred I. *Explaining Behavior: Reasons in a World of Causes*. Cambridge University Press, 1988.

Dymond, Rosalind F. “A Preliminary Investigation of the Relation of Insight and Empathy.” *Journal of consulting psychology* 12.4 (1948): 228. Print.

Eisenberg, Nancy. *The Development of Prosocial Behavior*. Academic Press, 1982. Print.

Ekman, Paul. “Facial Expressions of Emotion: New Findings, New Questions.” *Psychological science* 3.1 (1992): 34–38. Print.

Enç, Berent, and Fred Adams. “Functions and Goal Directedness.” *Philosophy of Science* (1992): 635–654. Print.

Feldstein, JoAnn Cohen, and Gerald A. Gladstein. “A Comparison of the Construct Validities of Four Measures of Empathy.” *Measurement and Evaluation in Guidance* 13.1 (1980): 49–57. Print.

Feshbach, Norma D., and Kiki Roe. “Empathy in Six- and Seven-Year-Olds.” *Child Development* 39.1 (1968): 133. *CrossRef*. Web.

Fiedler, Fred E. “A Comparison of Therapeutic Relationships in Psychoanalytic, Nondirective and Adlerian Therapy.” *Journal of consulting psychology* 14.6 (1950): 436. Print.

Field, Tiffany, Lisa Guy, and Vivian Umbel. “Infants’ Responses to Mothers’ Imitative Behaviors.” *Infant Mental Health Journal* 6.1 (1985): 40–44. Print.

Field, Tiffany M. et al. “Discrimination and Imitation of Facial Expressions by Neonates.” *Annual Progress in Child Psychiatry & Child Development* 16 (1982): 119–125. Print.

Frith, Uta. “Does the Autistic Child Have a ‘Theory of Mind’?.” *Cognition* 21 (1985): 37–46. Print.

Gallagher, Shaun. “Direct Perception in the Intersubjective Context.” *Consciousness and Cognition* 17.2 (2008): 535–543. *CrossRef*. Web.

---. “The Practice of Mind. Theory, Simulation or Primary Interaction?” *Journal of Consciousness Studies* 8.5-7 (2001): 5–7. Print.

Gallese, Vittorio, and Alvin Goldman. “Mirror Neurons and the Simulation Theory of Mind-Reading.” *Trends in cognitive sciences* 2.12 (1998): 493–501. Print.

---. “Mirror Neurons and the Simulation Theory of Mind-Reading.” *Trends in cognitive sciences* 2.12 (1998): 493–501. Print.

Gallese, V, C Keysers, and G Rizzolatti. “A Unifying View of the Basis of Social Cognition.” *Trends in Cognitive Sciences* 8.9 (2004): 396–403. *CrossRef*. Web.

Gecas, Viktor. “The Ebb and Flow of Sociological Interest in Values.” *Sociological Forum* 23.2 (2008): 344–350. *CrossRef*. Web.

George, Carol, and Mary Main. “Social Interactions of Young Abused Children: Approach, Avoidance, and Aggression.” *Child development* (1979): 306–318. Print.

Gilligan, Carol. *In a Different Voice: Psychological Theory and Women’s Development*. Cambridge, Mass.: Harvard University Press, 1982. Print.

Goldman, Alvin. “Two Routes to Empathy.” *Empathy: Philosophical and psychological perspectives* (2011): 31. Print.

Goldman, Alvin I. “Empathy, Mind, and Morals.” *Proceedings and Addresses of the American Philosophical Association* 66.3 (1992): 17. *CrossRef*. Web.

---. “In Defense of the Simulation Theory.” *Mind & Language* 7.1-2 (1992): 104–119. Print.

---. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, 2006.

Goldman, Alvin I. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford; New York: Oxford University Press, 2006. Print.

Gopnik, Alison, and Henry M. Wellman. “Why the Child’s Theory of Mind Really Is a Theory.” *Mind & Language* 7.1-2 (1992): 145–171. Print.

Graham, Loren R. "Between science and values." (1981).

Hall, Judith A., and Frank J. Bernieri. *Interpersonal Sensitivity: Theory and Measurement*. Psychology Press, 2001.

Heal, Jane. "Simulation, Theory, and Content." *Theories of theories of mind* (1996): 75–89. Print.

Hickok, Gregory. "Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans." *Journal of cognitive neuroscience* 21.7 (2009): 1229–1243. Print.

Hoffman, Martin L. "Development of prosocial motivation: Empathy and guilt." The development of prosocial behavior 281 (1982): 313.

---. "Developmental Synthesis of Affect and Cognition and Its Implications for Altruistic Motivation." *Developmental Psychology* 11.5 (1975): 607. Print.

---. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge University Press, 2000.

---. "Is Altruism Part of Human Nature?" *Journal of Personality and social Psychology* 40.1 (1981): 121. Print.

Iacoboni, Marco. "Mirroring People." *Farrar, Straus & Giroux, New York* (2008): Print.

Ickes, William. "Empathic Accuracy." *Journal of personality* 61.4 (1993): 587–610. Print.

Ickes, William, Victor Bissonnette, et al. "Implementing and Using the Dyadic Interaction Paradigm." (1990).

Ickes, William, Linda Stinson, et al. "Naturalistic Social Cognition: Empathic Accuracy in Mixed-Sex Dyads." *Journal of Personality and Social Psychology* 59.4 (1990): 730. Print.

Ickes, William, Eric Robertson, et al. "Naturalistic Social Cognition: Methodology, Assessment, and Validation." *Journal of Personality and Social Psychology* 51.1 (1986): 66. Print.

Ickes, William, Jeffry A. Simpson, and English Epigram. "Managing Empathic Accuracy in Close Relationships." *Close relationships: Key readings* (2004): 345–363. Print.

Kilner, J. M., and R. N. Lemon. "What We Know Currently about Mirror Neurons." *Current Biology* 23.23 (2013): R1057–R1062. Print.

Knafo, Ariel et al. "The Developmental Origins of a Disposition toward Empathy: Genetic and Environmental Contributions." *Emotion* 8.6 (2008): 737. Print.

Kuhlman, D. Michael, Curt R. Camac, and Denise A. Cunha. “Individual Differences in Social Orientation.” *Experimental social dilemmas* 3 (1986): 151–176. Print.

Kuhlman, D. Michael, and Alfred F. Marshello. “Individual Differences in Game Motivation as Moderators of Preprogrammed Strategy Effects in Prisoner’s Dilemma.” *Journal of personality and social psychology* 32.5 (1975): 922. Print.

Lamm, Claus, C. d Batson, and Jean Decety. “The Neural Substrate of Human Empathy: Effects of Perspective-Taking and Cognitive Appraisal.” *Cognitive Neuroscience, Journal of* 19.1 (2007): 42–58. Print.

Lazarus, Richard S. *Emotion and Adaptation*. New York: Oxford University Press, 1991. *Open WorldCat*. Web. 17 Dec. 2014.

Lovejoy, Arthur O. “Terminal and Adjectival Values.” *The Journal of Philosophy* 47.21 (1950): 593. *CrossRef*. Web.

McClintock, Charles G. “Social motivation—A Set of Propositions.” *Behavioral Science* 17.5 (1972): 438–454. Print.

McClintock, Charles G., and Wim B. Liebrand. “Role of Interdependence Structure, Individual Value Orientation, and Another’s Strategy in Social Decision Making: A Transformational Analysis.” *Journal of Personality and Social Psychology* 55.3 (1988): 396. Print.

McLaughlin, Barry. “Values in Behavioral Science.” *Journal of Religion and Health* 4.3 (1965): 258–279. Print.

McMullin, Ernan. “Values in Science.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*. JSTOR, 1982. 3–28.

Meltzoff, Andrew N. “Infant Imitation after a 1-Week Delay: Long-Term Memory for Novel Acts and Multiple Stimuli.” *Developmental Psychology* 24.4 (1988): 470. Print.

Meltzoff, Andrew N., and M. Keith Moore. “Imitation of Facial and Manual Gestures by Human Neonates.” *Brain Res* 19 (1974): 124. Print.

---. “Newborn Infants Imitate Adult Facial Gestures.” *Child Development* 54.3 (1983): 702. *CrossRef*. Web.

Meltzoff, Andrew N., and M. Keith Moore, others. “Imitation of Facial and Manual Gestures by Human Neonates.” *Science* 198.4312 (1977): 75–78. Print.

Morris, Charles. “Varieties of Human Value.” Midway Reprints, 1956.

Morris, Richard T. “A Typology of Norms.” *American Sociological Review* (1956): 610–613. Print.

Nichols, Shaun, and Stephen P. Stich. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds.* Clarendon Press/Oxford University Press, 2003.

Noddings, Nel. *Caring: A Relational Approach to Ethics and Moral Education*. University of California Press, 2003.

Panksepp, Jaak. "The Neuro-Evolutionary Cusp between Emotions and Cognitions: Implications for Understanding Consciousness and the Emergence of a Unified Mind Science." *Consciousness & Emotion* 1.1 (2001): 15–54. Print.

Pinker, Steven. *The Better Angels of Our Nature: The Decline of Violence in History and Its Causes*. Penguin UK, 2011. Print.

Preston, Stephanie D., and Frans De Waal. "Empathy: Its Ultimate and Proximate Bases." *Behavioral and brain sciences* 25.01 (2002): 1–20. Print.

Prinz, Jesse. "Against Empathy." *The Southern Journal of Philosophy* 49 (2011): 214–233. *CrossRef*. Web.

---. *The Emotional Construction of Morals*. Oxford University Press, 2007.

Prinz, Jesse J. *Gut Reactions a Perceptual Theory of Emotion*. Oxford; New York: Oxford University Press, 2004. *Open WorldCat*. Web. 17 Dec. 2014.

---. "Is Empathy Necessary for Morality?" *Empathy: Philosophical and psychological perspectives* (2011): 211–229. Print.

Rogers, Carl R. "The Necessary and Sufficient Conditions of Therapeutic Personality Change." *Psychotherapy: Theory, Research, Practice, Training* 44.3 (2007): 240–248. *CrossRef*. Web.

Rokeach, Milton, others. *The Nature of Human Values*. Vol. 438. Free press New York, 1973.

Ruby, Perrine, and Jean Decety. "How Would You Feel versus How Do You Think She Would Feel? A Neuroimaging Study of Perspective-Taking with Social Emotions." *Cognitive Neuroscience, Journal of* 16.6 (2004): 988–999. Print.

Sagi, Abraham, and Martin L. Hoffman. "Empathic Distress in the Newborn." *Developmental Psychology* 12.2 (1976): 175. Print.

Sawin, Douglas B. "Assessing Empathy in Children: A Search for an Elusive Construct." (1979): Print.

Selby-Bigge, L. A., and P. H. Nidditch. *David Hume: A Treatise of Human Nature*. Oxford: Oxford University Press, 1739. Print.

Simner, Marvin L. “Newborn’s Response to the Cry of Another Infant.” *Developmental Psychology* 5.1 (1971): 136. Print.

Singer, Tania. “The Neuronal Basis and Ontogeny of Empathy and Mind Reading: Review of Literature and Implications for Future Research.” *Neuroscience & Biobehavioral Reviews* 30.6 (2006): 855–863. Print.

Slovic, Paul. “If I Look at the Mass I Will Never Act”: Psychic Numbing and Genocide.” *Judgment and decision making* 2.2 (2007): 79–95. Print.

Sober, Elliott., and David Sloan. Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, Mass.: Harvard University Press, 1998. Print.

Sorensen, Roy A. “Self-Strengthening Empathy.” *Philosophy and Phenomenological Research* 58.1 (1998): 75. *CrossRef*. Web.

Stern, Daniel N. *The First Relationship: Infant and Mother*. Harvard University Press, 1977.

Stinson, Linda, and William Ickes. “Empathic Accuracy in the Interactions of Male Friends versus Male Strangers.” *Journal of personality and social psychology* 62.5 (1992): 787. Print.

Stotland, Ezra. “Exploratory Investigations of Empathy.” *Advances in experimental social psychology* 4 (1969): 271–314. Print.

Tetlock, Philip E. “Thinking the Unthinkable: Sacred Values and Taboo Cognitions.” *Trends in Cognitive Sciences* 7.7 (2003): 320–324. *CrossRef*. Web.

Van Overwalle, Frank, and Kris Baetens. "Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis." *Neuroimage* 48.3 (2009): 564-584.

Williams, Jr., Robin M., and Ethel M. Albert. “Values.” *International Encyclopedia of the Social Sciences*. Vol. 16. United States of America: Crowell Collier and Macmillan, 1968. Print.

Zahavi, Dan. “Beyond Empathy. Phenomenological Approaches to Intersubjectivity.” *Journal of Consciousness Studies* 8.5-7 (2001): 5–7. Print.

---. “Empathy and Direct Social Perception: A Phenomenological Proposal.” *Review of Philosophy and Psychology* 2.3 (2011): 541–558. *CrossRef*. Web.

---. “Empathy and Mirroring: Husserl and Gallese.” *Life, Subjectivity & Art*. Springer, 2012. 217–254.

---. “Simulation, Projection and Empathy.” *Consciousness and Cognition* 17.2 (2008): 514–522. Print.

Zahn-Waxler, Carolyn, JoAnn L. Robinson, and Robert N. Emde. “The Development of Empathy in Twins.” *Developmental psychology* 28.6 (1992): 1038. Print.

3. Two challenges to empathy from unconscious emotions

3.1 The two challenges

Theories of empathy can be divided into two account types: matching-accounts, and concern accounts. These types differ in terms of their criteria for *empathic accuracy*—the point at which the process of empathy ends. Whereas *matching accounts* focus on the experiential or representational match between an agent and a target, *concern accounts* focus on a match between an agent and a target's experienced states and an agent's awareness of a target's concerns. Both of these account types describe in detail the matching relationship required between an agent and target's *conscious emotional states* for accurate empathy to occur. But recently, developments in psychoanalysis, evolutionary psychology, and embodied cognition have made it clear that *unconscious emotional states* play an important role in our mental lives. The existence of unconscious emotional states has important implications that have not been investigated for what counts as accurate empathy on each type of empathy account. This paper is about two challenges that unconscious emotions present for accurate empathy on existing theories of empathy. First I reintroduce the distinction from the previous paper between matching and concern accounts of empathy. Then, I describe what I take unconscious emotions to be, and what accurate empathy is on each account type. I go on to present a challenge to each account type from unconscious emotions. The first of these is the *unconscious matching challenge*. The second is the *unconscious concern challenge*. I address each of these in turn. I present replies from the perspective of a theorist espousing a matching account of empathy to the first challenge, and a reply from the perspective of a theorist espousing a concern account of empathy to the second challenge. The reason for presenting my replies in this way is that each challenge appeals to the matching relationship between an agent and a target that is appropriate to each account type. The later reply draws from Charland's (2005) *indeterminacy thesis of affect valence*. It shows that unconscious emotions are not functionally and behaviourally identical to conscious emotions, and hence should not be treated as such on concern accounts of empathy. It will turn out that there are a variety of possible responses to these challenges from

unconscious emotions, but that a researcher espousing either account type of empathy can meet them successfully.

3.2 Two types of accounts: *matching* and *concern*

In the previous paper I distinguished between two types of empathy accounts: *matching-accounts* and *concern-accounts*. The account types differ in how they treat *empathic accuracy*. In this section I will briefly review the notion of empathic accuracy and how these account types are differentiated by their criteria of empathic accuracy.

As the name suggests, "empathic accuracy" is the phenomenon whereby an agent accurately empathizes with a target. It does not refer to the process of empathy as a whole. Rather, it refers to the *outcome* of empathy (Ickes 1993). As such, it does not prioritize any of the component processes of empathy (e.g. neural mimicry, emotional contagion). Empathic accuracy can be achieved by varying routes. Goldman (2011) argues that empathic accuracy can be achieved by two routes: the "mirroring route" and the "reconstructive route". The *mirroring route* is involuntary, and it involves the component process of mirror neuron activation, whereas the *reconstructive route* is voluntary and it involves the component process of perspective-taking.

Individual accounts of empathy emphasize different routes to accurate empathy. As a consequence of this, they also emphasize different component processes of empathy. But I have argued that accounts of empathy can be classified as tokens of one of two types according to their criteria for empathic accuracy: *matching accounts* (Ickes 1993; Gallese and Goldman 1998; Goldman 2006, 2011; Iacaboni 2008; de Vignemont and Singer 2006; Prinz 2011a, 2011b) and *concern accounts* (Krebs 1975; Batson 1991; Sober & Wilson 1998; Hoffman 2011).

On a matching account, empathic accuracy occurs when an agent's internal state *matches* the internal state of a target. This matching relation can take one of two forms:

> **Matching relation A:** An agent has the same experiential states as those of a target

or

> **Matching relation B:** An agent *infers* (or represents) the content of a target's experiential states

On matching accounts, accurate empathy occurs when one of these two forms of matching between agent and target occur. For example, imagine a runner participating in a race that is running in second place. This runner is, at this particular time, feeling hopeful because they expect to soon pass the first place runner. On a matching account. For an agent to accurately empathize with this target runner, the agent will either feel hopeful themselves or think that the target is hopeful.

On concern accounts, accurate empathy occurs when an agent becomes aware of a target's concerns and feels motivated to help that target with their concerns. Like on matching accounts, concern accounts also involve a matching relation between an agent and a target. But it is important to note that the matching relation on concern accounts is not the same as either of the two forms of matching relations just mentioned above. On concern-accounts, the matching relation between an agent and a target is that of matched *affect valence*. Affects are individual felt emotional states. They can be classified as either positive or negative. For example, if I am sad at the loss of a loved one, I would classify the valence of this emotion as negative. If I am proud that my friend won a race, I would classify the valence of this emotion as positive. The classification of emotional states as either positive or negative is to attribute to them a *valence* (Lazarus 1991). Accordingly, accurate empathy on concern accounts take the following form:

> **Matching relation C:** The valences of the agent's experience match the valences of the target's experience.

Concern accounts also add that, when accurate empathy occurs, an agent feels motivated (at least momentarily) to help a target with their concerns. On concern accounts, when *matching relation C* occurs it leads an agent to become aware of a target's concerns. To

understand how this occurs, let us return to the example of the runner participating in a race. For accurate empathy to occur in this case on concern accounts, an agent must first match the emotional state valences of the target. So when the runner wins the race and feels pleasure or joy, the runner, in this example, is experiencing an emotional state with a positive valence. The agent need not match the *content* of the target's emotional states. That is, the agent does not need to feel pleasure or joy. Rather, what is required is that the agent experience emotional states with the same valences. So an agent that is accurately empathizing with the runner may be experiencing pride or relief (whereas the runner is experiencing pleasure and joy). This match of valence then leads the agent to become aware of the target's concerns. These concerns may include: the runner's concern to share their joy with the other runners; the runner's concern to be recognized for their efforts in winning the difficult race, and so on. Having become aware of the runner's concerns, the agent will then feel motivated to help the runner with those concerns. It is *not* required the agent actually behave in a helpful way. However, what is required is that the agent *feel* motivated to help the target with their concerns. Accurate empathy occurs on a concern account when there is an emotional valence match, and the agent feels motivated to help the target with their concerns.

On both matching-accounts and concern-accounts, an emotional matching relation between agent and target is required for empathic accuracy. On a *matching account*, accurate empathy requires that an agent match the *content* of the target's emotional states. And on a *concern accounts*, accurate empathy requires that an agent match the *valences* of the target's emotional states. This accords with the well-established fact that empathy minimally requires a match between an agent and a target's *conscious* emotional states. But the role of *unconscious emotions* in empathic accuracy remains unexamined. I will present a challenge from unconscious emotions to each empathy account type. This will allow me to specify the implications that unconscious emotions carry for accounts of empathy. But before presenting these two challenges, I will first specify what *unconscious emotions* are.

3.3 Unconscious emotions

By introspection we can tell that emotions are felt; they involve conscious feelings (James 1884). During the early part of the twentieth century researchers (particularly in philosophy) focused their efforts on developing theories of conscious processes. But recently, a growing number of researchers in philosophy, psychology, and psychiatry have claimed that *unconscious emotions* are just as real as conscious ones (Rosenthal 1991; Berridge & Winkielman 2003; Prinz 2004; Wilson 2004; Winkielman et al. 2005). These researchers were spurred by the development of Freud's "new science" of psychoanalysis. Freud posited that certain mental processes occurred independently of consciousness (Freud 1895). Since then, developments in neurophysiology (Woody and Phillips 1995), the informational turn in philosophy of the 1950's (Adams 2003), and the cognitive revolution in psychology of the 60's and 70's (Westen 2002) have documented that much psychological processing is independent of conscious experience or awareness. This has been especially relevant to researchers working on emotions because the existence of unconscious mental processes allows for possibility of *unconscious emotions* (Wilson 2003).

Researchers now generally take unconscious emotions to be states of an agent that are identical to conscious emotional states, except that they are not felt or reportable while they are instantiated (Kihlstrom 1995, 1999; Kihlstrom et al. 2000; Prinz 2004). Three areas of study have been particularly influential in promoting this view: psychoanalysis (as mentioned), evolutionary psychology, and embodied cognition. For example, evolutionary psychologists claim that humans have evolved a system for recognizing and reacting to emotional stimuli independent of consciousness (Lundqvist and Öhman 2005). They argue that threatening stimuli, such as angry facial expressions, unconsciously activate subcortical sites in the brain, and rapidly redirect attention before the stimuli have been consciously experienced (Lundqvist and Öhman 2005, p 105). What these psychologists are saying is that felt states such as fear extend beyond conscious awareness. Further, they hold that the functional and behavioural contribution of these unconscious precursors to emotional states can occur independently of conscious

experience altogether. Accordingly, we can be fearful without feeling fear (Dimberg et al. 2000).

The second area of study in which unconscious emotions feature prominently is in *embodied cognition*. Here psychologists have claimed that unconscious emotions influence both preferences and behaviour (Winkielman et al. 2005; Winkielman et al. 2008). In many studies, participants have been exposed to photos of happy and angry faces, and asked to sort those photos according to criteria independent of the subject in the photo's facial expression. Then participants were asked to pour and drink beverages they had not drank before. It was found that being exposed to more happy faces than angry faces resulted in participants pouring and drinking more of the beverages. The interpretation of these results is that the unconscious emotion of happiness biases the processing of subsequent stimuli and guides subsequent behaviour (Semin 2008).

Other research in embodied cognition suggests that unconscious emotional processing influences the speed of sorting tasks. In the experiments of Neumann and Strack (2000), participants were asked to sort positive, neutral, and negative words. As the words were presented to participants on a computer screen, circles appeared behind the words as either increasing in size (moving toward the participants) or decreasing in size (moving away from the participant). It was found that participants sorted negative terms more quickly when the circles were moving away from the participants, and positive terms were sorted more quickly when the circles were moving toward the participants. These results are taken to show that emotional states can be influenced unconsciously by visual cues of approach and avoidance (*Ibid.*). They lend credence to the claim that unconscious emotional processing affects consciously controlled behaviour.

A final example is that of Freudian psychoanalytic theory. For Freud, unconscious states play an important role. He argues that emotional states can be unconscious in at least two ways. The first of these is when they are *repressed* (Freud, 1915). An unconscious emotion that is repressed is one that, at a previous point, may have been experienced consciously. But because this emotion was negative, the agent has *defended* against it by associating it with an idea separate from the original emotion. Thus, the emotion is no

longer experienced consciously. The agent is only conscious of the idea now associated with the previously experienced emotion. The experience of the associated emotion has been repressed.

Another way that unconscious emotions feature in Freudian psychoanalysis is in the phenomenon of *displacement* (Freud, 1926). This is a phenomenon whereby one emotion (e.g. hate) is directed at an intentional object consciously in way that differs from the emotion's unconscious intentional object. Here Freud provides the example of a boy who develops a hatred for horses consciously, but is in truly expressing an unconscious hatred for his father. When the boy sees his horse get injured, he wishes that it were his father instead of the horse who is hurt. Rather than accepting this fear of his father, the boy comes to replace fear with hatred for horses. The consciously felt emotion of fear develops into hatred, and is then displaced or misdirected at horses. The hatred that the boy has for horses is thus said to be displaced *unconscious* hatred for his father (*Ibid.*).

In the three research areas mentioned above (evolutionary psychology, embodied cognition, and psychoanalysis) unconscious emotions are states that play the same functional and behavioural roles as conscious ones, except that they are not felt or reportable when they occur. Nonetheless, they make the same functional and motivational contributions as conscious emotions (Erdelyi 1985; Kihlstrom et al. 1984; Bargh et al. 1982; Bargh et al. 2001; Bargh & Morsella 2008; Bargh and Morsella 2010). The question then is what requirements do unconscious emotions place on the two types of empathy accounts and empathic accuracy? The two challenges from unconscious emotions and the replies to these challenges I will present allow me to specify the implications that unconscious emotions carry for accounts of empathy.

3.4 Empathy in three parts

Before presenting my two challenges to accounts of empathy, I will in this section review what empathy is. My description in three parts will allow me to fix reference on *accurate empathy*. It is important to distinguish *accurate empathy* from the whole process of *empathy* because accurate empathy refers to the end state of a successful attempt at empathy. Empathy can be broadly divided into three parts:

1) Initial conditions

2) The processes that instantiate it (component processes)

3) The end state (empathic accuracy)

A description of empathy's initial conditions (1) will include the relationship between an agent of empathy, a potential target of empathy, and an environmental context in which the agent and the target interact. Which targets are selected by an agent as potential targets of empathy is a complex issue that I have addressed in the previous paper. I will not address it here. For present purposes, I merely wish to distinguish the initial conditions of empathy from the ongoing activity of its component processes. So, once empathy gets started it can follow one of two procedural routes: the *involuntary route* and the *voluntary route*.[32] On the involuntary route, an agent perceives a target, and the processes of empathy are activated automatically. On the voluntary route, an agent voluntarily chooses a potential target of empathy that is either perceived, has been perceived, or is imagined.

The component psychological processes that instantiate empathy can differ on each of these routes. For example, on the involuntary route, I may come across a target asking for money on the street. Upon perceiving this target, the process of *emotional contagion* may be activated automatically. Emotional contagion is a process by which an agent involuntarily "catches" the emotions of a target (Doherty 1997). Psychological experiments have shown that the capacity for this response is either innate or that it can occur very early in development. Two day old infants cry when they hear another newborn cry (Simner 1971; Sagi & Hoffman 1976). And it has been shown that this response is only stimulated by the crying of other infants as opposed to other loud sounds (*Ibid.*). In adults too, emotional contagion is a "rudimentary empathic distress reaction" whereby an agent experiences a vicarious emotional response to the emotions expressed by a target (Knafo et al. 2008 p 3; Zahn-Waxler et al. 1992). Returning to our example then, the target asking for money on the street may appear cold and sad. Accordingly,

[32] Goldman (2011) calls these the "mirroring route" and the "reconstructive route".

when empathy gets started, I will "catch" this target's emotional states involuntarily. These processes will cause me to experience feelings of sadness. Other processes too may instantiate empathy on the involuntary route. These include *motor mimicry* and *neural mimicry*. Because I have discussed these processes in detail in the previous paper, I will not do so here. What is important to emphasize for present purposes is that, once empathy gets started, there are many involuntary processes that lead to empathic accuracy.

As mentioned, on the involuntary route the component process of empathy can differ from those on that instantiate empathy on the voluntary route. Under voluntary control are empathy's component processes of *like-me perspective-taking* and *like-other perspective-taking*. For example, when I converse with a friend, and that friend tells me that they have recently lost their job, I may decide to "take their perspective". When engaging in *like-me perspective-taking*, I imagine what it would be like to be in my friend's situation *as myself* (as the agent). I will imagine *myself* in that target's situation. I will imagine what the effects of losing my job would be on me. So for example, if I would be saddened by losing my job, I will imagine that the target–having lost their job–will feel sad. In like-me perspective-taking, the agent uses their familiarity with situations like that of the target to imagine how they (as this agent) would think and feel were they in the target's situation.

Alternatively on the voluntary route, I may engage in like-other perspective-taking. Like-other perspective-taking differs from like-me perspective-taking in that there is a voluntary effort on the part of the agent to concentrate on what it would be like for the target (not the agent themselves) to be in that situation. Like-other perspective-taking is a process in which an agent takes into consideration beliefs about the particular target. These beliefs may include beliefs about a target's character, their present context, their history, their values, and so on. What differentiates this mode of perspective-taking is that I imagine how the particular target's situation is affecting them, rather than relying on my familiarity with a type of situation the target is in and my prototypical beliefs about what I would experience were I in the target's situation.

The processes along the *voluntary* route to empathy can, and often do, recruit the *involuntary* processes typical of the involuntary route. For example, when an agent takes the perspective of a target, this agent may also be sharing some of the agent's emotional state by means of emotional contagion or motor mimicry. Alternatively, when perspective-taking, an agent's mirror-neuron system may be activated. So the component processes of empathy can be treated as overlapping along the voluntary route. Similarly, involuntary component processes can lead an agent to activate voluntary ones. For example, if a target is selected as a potential target of empathy involuntarily by the activation of emotional contagion, this process may lead an agent to voluntarily take that target's perspective in attempt to more accurately empathize with that target.

We have just seen that once empathy gets started, it can proceed along two routes. Along these routes, the component processes of empathy will be activated. And eventually, they will stop. It is when empathy stops that we can carve out a portion of the phenomenon to be assessed as accurate or inaccurate. This is the outcome of empathy. I have described above that the part of empathy that is counted as empathic accuracy will differ on each account type (matching or concern). On a matching account, accurate empathy occurs when either *matching relation (A)* or *matching relation (B)* are instantiated. These matching relations can be combined as follows:

> **Accurate empathy on a *matching* account:** when an agent experiences [matching relation A] or represents [matching relation B] the content of a target's emotional state.

Which of these relations (A or B) counts as accurate empathy on a matching account will vary from one token of this type to another. But what all matching accounts share is that accurate empathy occurs when there is a match between an agent and the exact content of a target's emotional state.

As we have also seen above, when empathy stops on a concern account, accurate empathy occurs when *matching relation (C)* is instantiated. On this account type, accurate empathy occurs when an agent matches the valences of a target's emotional states. There need not be a match between the content of the agent and the target's

experiential states. But on concern accounts this is not all that accurate empathy circumscribes. The agent also needs to be aware of the target's concerns and feel motivated to help the target with those concerns. These additional conditions are introduced on a concern account to better explain the relationship between accurate empathy and helping behaviour. For example, when speaking to a friend who has just lost their job, accurate empathy will occur when the agent matches the *valences* (as opposed to the *content*) or a target's emotional states. If the target is feeling sad, the agent may be feeling disappointed. So far so good, as long as both the agent's and target's emotional states are congruently valenced—in this example, *negatively*. But to repeat, concern accounts add two criteria of empathic accuracy. The agent must become aware of the target's concerns such as: communicating their negative emotions with a friend; feeling re-assured that they will find another job; finding another a job, and so on. And an agent must feel *motivated* to help the target with their concerns. Thus, accurate empathy on a concern account can be summarized as follows:

> **Accurate empathy on a *concern* account:** (1) when an agent experiences a congruently valenced (with a target) emotional state; (2) is aware of the target's concerns; (3) feels motivated to help the target with their concerns.

We can now see how matching accounts and concern accounts of empathy each employ a different notion of empathic accuracy. I will now present the *first challenge from unconscious emotions* to each account type of empathy.

3.5 Challenge I: Unconscious matching

As we have seen, an account of empathy can employ one of three matching relations to determine when accurate empathy occurs. In doing so, accounts have focused on the matching relation between an agent and a target's *conscious emotions*. Conscious emotions are felt emotional states such as sadness or anger. But such emotions can occur independently of *awareness* or *conscious experience* (Rosenthal 1991). The question then is what requirements do unconscious emotions place on each account type's notion of empathic accuracy?

The first challenge from unconscious emotions is the *unconscious matching challenge*. This challenge is directed at matching accounts of empathy. We have seen that on matching accounts, two types of emotional matching are possible for accurate empathy to occur. An agent may *experience* the same conscious emotional states as a target. Or an agent may *represent* (without necessarily experiencing) the content of a target's conscious emotional states. When combined with a realist position on unconscious emotions, we can begin to ask questions like: What sort of matching relation would be required if an account of empathy integrated the role of unconscious emotions?

Unconscious emotions are either unfelt or unreportable while they are instantiated, but they share two features with conscious emotions. First, unconscious emotions are *functionally equivalent* to conscious emotions. This means that they play the same roles as conscious emotions within an agent's overall psychology. So for example, an agent who is consciously happy will form preferences for new beverages more easily than an agent that is consciously sad. Similarly, if an agent is primed to be unconsciously happy (or to be unaware of their happiness), this agent too will form a liking for new beverages more easily. The second feature that unconscious emotions share with conscious ones is that they are taken to be potentially *behaviourally equivalent*. This means that the causal contribution of an emotional state to an agent's behaviour can be the same regardless of whether the state is conscious or unconscious. Returning to a previous example, an agent may feel motivated to harm his horse and do so because he is consciously angry at the horse. Alternatively, an agent may harm his horse because that agent has displaced fear of his father that is now being directed at his horse. In the former case, the conscious emotion of anger results in the agent harming his horse. And in the latter case, the unconsciously displaced emotion of fear is resulting in the same behavioural outcome. On a matching account, accurate empathy occurs when an agent matches the *conscious* emotional states of a target (Darwall 1998; Prinz 2011a, 2011b). The unconscious matching challenge then asks of matching accounts whether an agent must match both the target's conscious *and* unconscious emotional states; and if so, how is this achieved.

Including unconscious emotions when formulating the matching relations that are possible on matching accounts of empathy allows us to see the problem more clearly. If

on a matching account, an agent must match the unconscious emotions of a target, two possible matching relations are possible:

> **Matching relation A_2:** An agent has the same *unconscious emotions* as that of a target

or

> **Matching relation B_2:** An agent represents the content of a target's *unconscious emotions*

I think that we can meet the *unconscious matching challenge* on a matching account that employs either of the above matching relations for accurate empathy. I will present three replies to this challenge. Our aim in trying to solve this puzzle is to determine whether matching accounts of empathy can retain a useful notion of empathic accuracy. At stake is our ability to assess empathic accuracy on matching accounts. On such accounts, accurate empathy requires that the agent and that target's states be identical. This may be very difficult to assess if it requires that both the agents and targets have the same unconscious emotional states. Furthermore, even if the difficulty of assessing whether accurate empathy has occurred is overcome, the unconscious matching challenge may make it difficult to defend a matching account of empathy if empathy occurs only very rarely on such an account.

3.5.1 First reply: Eliminativism

'Eliminativism' is a term used to refer to the thesis that a category of entities, processes, or properties do not exist (Ramsey et al. 1990). A standard example is eliminativism about witches. In the past, witches were posited in explanations of various calamities (*Ibid.*). But they have since been eliminated from explanations of these same calamities. People concluded that positing witches did not contribute to explaining the phenomena under consideration. Repeatedly in science, certain entities are posited and then later eliminated in subsequent theories. For example, phlogiston was posited to explain combustion and rusting. Items were said to contain phlogiston that was released when burnt. This release explained the burning process and why items lost mass after burning.

Phlogiston was later eliminated from these theories in light of evidence showing that certain substances (such as magnesium) actually gained in mass when burned. There is evidence that information processing occurs unconsciously, and some compelling evidence of their functional contribution to decision-making in embodied cognition. I will not attempt to argue for whether unconscious emotions should be eliminated on a matching account of empathy. I present the position merely as one that can meet the *unconscious matching challenge.*

A researcher proposing a matching account of empathy can meet the unconscious matching challenge by being an *eliminativist* about unconscious emotions. The researcher can accept that much "information processing" goes on unconsciously. For example, many component processes of vision operate independent of consciousness. Evidence of this is borne out in cases of brain injury. Agents with brain lesions sometimes experience *hemi-neglect*, whereby they respond to visual stimuli in the visual field that is contralateral to the area of the lesion without consciously experiencing the stimuli (Gelder 2005). An empathy researcher may accept that this type of unconscious information processing occurs, but deny the existence of unconscious emotions for at least two reasons. First, it is difficult to identify and explain the connections between unconscious emotions, conscious states (including conscious emotions), and behaviours (Izard 2009, p 12). This is especially problematic in psychoanalysis where the identification of the roles that unconscious emotions play requires long-term interaction and familiarity with an agent's particular situation and history. Second, although these connections have been identified more robustly in evolutionary psychology and embodied cognition, most unconscious emotional states have yet to be correlated with neural states (*Ibid.*). Some of the psychological processing involved in facial recognition, for example, has been neurally correlated (Ekman 2003). However, many unconscious emotions taken to influence preferences (as in the beverage drinking experiment), and motivations (in evolutionary psychology) have not been correlated with neural states. Thus, there are reasons for being an eliminativist about unconscious emotions. It is difficult to explain their functional contribution, and it is difficult to identify their neural correlates.

The eliminitavist can answer "no" to the question of whether empathy requires that an agent match both a target's conscious and unconscious emotions. This is because, on this view, unconscious emotional states do not exist. If unconscious emotional states do not exist, a target of empathy does not have any such states. Therefore, an agent *cannot* match the unconscious emotions of a target.

3.5.2 Second reply: Restrictivism

'Restrictivism' is the term I will be using to refer to a theory that *restricts* an explanation of the phenomenon under consideration to only certain ontological posits. My usage of the term is close to that of Prinz (2011c) who calls "restrictivism" the position that all conscious states are perceptual states. Prinz argues that an explanation of cognitive phenomenology can be *restricted* to positing mental images and inner speech. The details of his position are not important for present purposes. What I would like to emphasize (and use) is the argumentative or theoretical strategy that Prinz employs. Restrictivism on this usage refers to a position which restricts the explanation of a phenomenon X to a restricted set of entities. The grounds for why an explanation should be restricted are not often made explicit. But the principle seems to involve the claim that the restricted theory can do everything the less restricted theory can do (if not better).[33] In Prinz's case, restrictivism involves explaining all conscious experience by appeal to a restricted set of entities, namely mental images and inner speech acts (*Ibid.*). This same argumentative strategy can be used by a researcher espousing a matching account of empathy to meet the unconscious matching challenge.

The unconscious matching challenges asks whether, on a matching account, an agent needs to match the unconscious emotions of a target for accurate empathy to occur. We have seen that on a matching account, two types of matching relations are possible:

[33] Restrictivism seems to be equivalent to Ockham's razor argument.

Matching relation A: An agent has the same experiences as those of a target

or

Matching relation B: An agent *infers* (or represents) the content of a target's experience

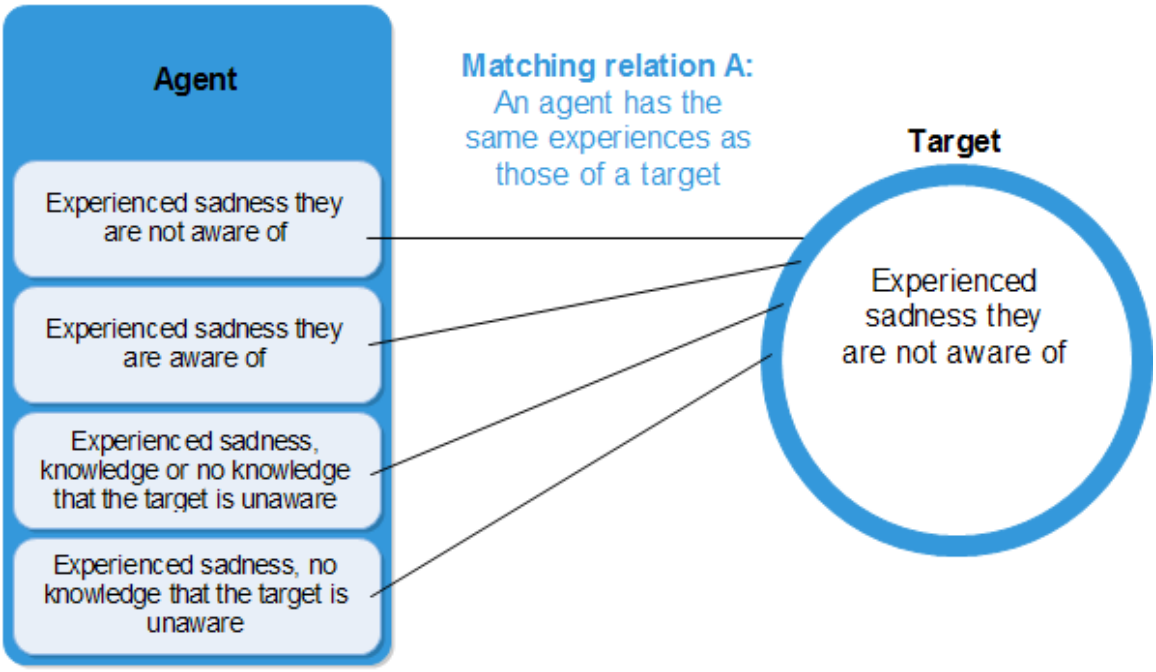
Restrictivism about unconscious emotions can be used to meet the unconscious matching challenge on an account where *matching relation A* is the criterion for accurate empathy. On *matching relation A*, accurate empathy occurs when *experienced* emotional states of an agent match those of a target (Iacaboni 2008; Prinz 2011a, 2011b). It follows that because a target's unconscious emotions are not felt experiences, an agent that this empathizing with this target cannot have matching experiences of such states. An agent need not *have* the same unconscious emotional states as a target (*matching relation A*$_2$) because an explanation of accurate empathy can be *restricted* to states that are consciously experienced. Thus, the unconscious matching challenge is met by restricting the ontology to be used in explanations of accurate empathy to a match between an agent and a target's consciously experienced states.

Before moving on, I will consider a variation on matching relation A that takes into consideration a different notion of unconscious emotions. In this section, I have treated unconscious emotions as unfelt states lacking in phenomenal experiential content. However, Rosenthal (1991) treats unconscious emotions differently. For him, unconscious emotions can be experienced, but they occur independently of *awareness*. For example, a target may experience anger but not recognize or be aware of the experience (*Ibid.*). But even when there is no recognition or awareness of this conscious emotional state, the target nonetheless is said to have an emotion with an experienced phenomenal quality. Thus, for Rosenthal, unconscious emotions are states that we feel but are not attentively aware of. How then would a researcher espousing an account of empathy with *matching relation A* and Rosenthal's notion of unconscious emotions respond to the unconscious matching challenge? Let us first look at what *matching relation A* requires of Rosenthal's notion of unconscious emotions.

Let us assume that a potential target of empathy is experiencing sadness that they are not aware of. *Matching relation A* would then require that an agent also experience this sadness. So far so good. Having restricted the explanation of accurate empathy to experienced states, a researcher can accommodate Rosenthal's notion of unconscious emotions on a matching account as long as the agent matches the target's unaware but felt states. However, it is unclear whether such an account of empathy requires that an agent merely experience the same unaware emotional states as the target, or whether the account also requires that the agent be unaware of their own experienced emotional states.

Alternatively, a less stringent account would not impose this condition on the agent's awareness. Accurate empathy would occur when the agent experiences the same unaware emotional states as a target regardless of whether the agent knows that the target is unaware of them. An even less stringent account would impose even less on the agent's emotional awareness. As in the previous variation, accurate empathy would occur when the agent experiences the same unaware emotional states as the target and the agent need not have any awareness of the target's awareness (or lack thereof) of their states. Furthermore, the agent themselves may not be aware that they are experiencing the emotional states being matched between agent and target.

Figure 3: *An agent matching a target's experienced emotions outside of awareness*



For Rosenthal, unconscious emotions are experienced states that we are not aware of. Adopting restrictivism can be used to meet the unconscious matching challenge on a standard matching account or a matching account that uses Rosenthal's notion of unconscious (unaware) emotions. What restrictivism requires is that only *experienced* states be matched between agent and target for accurate empathy to occur. It does not necessarily impose conditions on an agent's awareness of their own emotions, nor on an agent's beliefs about a target's emotional awareness. It merely requires that both agent and target have the same emotional experiences regardless of whether either the agent or the target are aware of those experiences.

3.5.3 Third reply: Rejecting the challenge

We have just seen that restrictivism is a reply to the unconscious matching challenge on a matching account of empathy that uses *matching relation A*. I turn now to a reply to this first challenge that is directed at matching accounts of empathy that use matching relation B, instead. When accurate empathy occurs on a matching account that employs matching relation B, an agent infers to (or represents) the content of a target's emotional experience. The question then is how can one reply to the unconscious matching

challenge on such an account of empathy? Does such an account require that an agent also represent the content of a target's unconscious emotional states for accurate empathy to occur? [34]

The simple reply to the unconscious matching challenge in this case is "yes". I have titled this reply to the unconscious matching challenge a "rejection" of the challenge because on a matching account that uses *matching relation B*, this challenge, in a sense, poses no challenge. A matching account that uses *matching relation B* can meet the unconscious matching challenge without amendment or qualification to the account. Accurate empathy will simply occur when an agent represents the content of a target's unconscious emotional states. For example, if a target has an unconscious emotional state of happiness (as in the new beverage drinking experiments), then an agent must *represent* that agent's unconscious emotional state for accurate empathy to occur. On this reply, a researcher need not adopt eliminativism about unconscious emotions. Nor does the researcher need to restrict the set of explanatory entities to those consciously experienced. All an agent needs to be able to do for accurate empathy to occur is have a mental representation about the content of a target's emotional experiences. However, a difficulty with this response is that it is unclear how an agent comes to represent the target's unconscious emotional states.

I called this "rejection reply" simple. It is a simple reply because it rejects the unconscious matching challenge by meeting it without change to the account. But its simplicity veils the complexities that arise from putting such an account of empathy into practice. The case of the beverage drinking experiment mentioned above was relatively straightforward. In that case, the agent was an experimenter, and the target was a participant in an experiment whose hypothesis was about the relationship between unconscious emotions and decisions to drink a beverage. In such a case, it is easy for the agent to hypothesize (and thus represent) the unconscious emotions that the target has because the agent has constructed the environment in which the agent and the target interact for the very purpose of proving this hypothesis. The agent (experimenter) has

[34] See matching relation B_2.

primed the unsuspecting target (participant) to have certain unconscious emotional states (e.g. unconscious happiness). Then the agent has inferred the relationship between these unconscious emotions and the behaviours of the target. Thus, the agent accurately empathizes with the target by forming an accurate representation of the target's unconscious emotional states. But an accurate match between an agent's representations and the content of a target's unconscious emotional states is not always so easily achieved. Even in this experimental case, we might be skeptical of the claim that the experiment has successfully primed unconscious happiness.

Recall that the first reply—eliminativism—was considered on the grounds that (1) unconscious emotions are difficult to identify and trace, and (2) that they are not easily correlated with brain states. If a researcher maintains that unconscious emotions exist while espousing a matching-account of empathy (with *matching relation B*), then that researcher must also provide an account of how accurate empathy occurs in less artificial environments of agent / target interaction. An example of such an environment is one where the agent is a psychoanalyst and the target is a client. In this case, the agent may have the theoretical knowledge to identify a target's unconscious emotional states and the skill to determine the connection between them and the target's behaviour. However, in the best of cases, this requires that the agent and the target interact over an extended time. The agent must have a detailed familiarity with the particular target and how the situations of their lives are affecting them. But this is not something that everyone is trained to do. As mentioned, specific education and skill are prerequisites for reliably and accurately representing the unconscious emotions of targets. And even then, there exists intense and ongoing theoretical disagreement between the various schools and sub-disciplines of psychoanalysis and psychiatry generally. How can an agent be sure that they are accurately representing the unconscious emotions of a target? Thus, the unconscious matching challenge presents a serious epistemological problem for matching accounts of empathy.

Matching accounts have not sufficiently explored the problem of how an agent is able to reliably form representations of a target's *unconscious* emotional states. Even experiments performed to show how agents represent a target's *conscious* emotional

states generally have not required that participants have any special training in identifying the unconscious emotions of targets. For example, Icke's "dyadic interaction paradigm" for measuring empathy is based on a matching account this type (Ickes & Tooke, 1988; Ickes et al. 1990; Stinson & Ickes, 1992).[35] Briefly, it involves two participants interacting for five minutes while unknowingly being videotaped. Then, the participants are asked to re-watch the video of their interaction several times noting at co-varying points what they (the agent) were thinking and feeling and what their interaction partner (the target) was thinking and feeling. When there is a match between what the agent takes the target to have been experiencing and what the target reports to have been experiencing, a point is assigned to the agent for having accurately empathized with the target. Accordingly, such a matching account can meet the unconscious matching challenge as long as the agent is also able to accurately represent the content of the target's *unconscious* emotional states.[36] But without the relevant training, it is unlikely that this will occur frequently. Even in current dyadic interaction experiments where agents are limited by an experimenter's instructions to inferring what targets are *consciously* thinking a feeling, their empathic abilities vary widely (*Ibid.*). It follows that it will be even more difficult for an agent to accurately represent a target's *unconscious* emotional states in environments (outside of a laboratory) where no such limits exist.

For accurate empathy to occur on a matching-account (with *matching relation B*), agents must accurately represent both the conscious and the unconscious emotions of their interactions partners. Although such an account can reject the unconscious matching challenge outright, this reveals that its criteria for accurate empathy are stringent. They place a high demand on the abilities of agents to represent the content of a target's unconscious states. And as we have seen, even in stable laboratory environments, such accurate representations are hard to come by.

---

[35] In some studies, this paradigm has been used to measure therapist's empathic ability (Ickes 1993).
[36] The target's unconscious emotional states would have to be independently identified because the target would not be conscious of them.

3.6 Challenge II: Unconscious concerns

I will now move on to present and reply to the second challenge from unconscious emotions: the unconscious concerns challenge. Whereas the first challenge from unconscious emotions was directed at matching accounts of empathy that employed matching relations A and B, this second challenge is directed at concern accounts of empathy that employ *matching relation C*. Recall that on matching accounts, accurate empathy occurs when an agent becomes aware of a target's concerns and feels motivated to help that target with their concerns. The agent's awareness of the target's concerns on concern accounts results from an affective match whose form is different from the matching relations we have seen used on matching accounts. Rather than a match in experiencing or representing a target's emotional states, concern accounts employ a form of matching whereby an agent matches the valences of the target's emotional states. On concern accounts, the *emotional valence-match* between agent and target contributes to the agent's awareness of the target's concerns, and accurate empathy occurs when this valence match causes an agent to feel motivated to help the target with their concerns (Batson and Shaw 1991; Hoffman 2012). But what if the *true concerns* of the target (as opposed to the apparent concern) are caused by unconscious emotions?

It is well established that *conscious* emotions are motivational (Maslow et al. 1954; Lazarus 1991; Prinz 2004). Many of our everyday and mundane *concerns* are motivated by conscious emotional states. For example, when someone insults us, we may feel anger. Our concern to retaliate against the offender will be motivated by this anger. Alternatively, when alone and encountering someone for the first time in an isolated location, we may feel fear. And this feeling of fear may motivate our concern to avoid or flee from the person we are encountering. Accordingly, we can say that emotions motivate concerns. I am using a 'concern' here to include things like *goals* and *needs.* A target may put more effort into work because they are concerned about losing their job. Or a target may be concerned about finding a new job after having looked for one for a long time. The concerns in these cases will be motivated by conscious emotions such as a joy or anxiety. But what if an agent's behaviours and concerns are motivated by unconscious emotional states? The challenge of unconscious concerns asks: must an

agent be aware of the unconscious concerns of a target to accurately empathize with that target? Clearly, if an agent misperceives the concerns of a target, the agent may still be motivated to alleviate those falsely perceived concerns. The challenge of unconscious needs does not deny this. Rather it asks: does accurate empathy on a concern account require that an agent become aware of the unconscious concerns of a target as opposed to the target's apparent concerns?

To better understand the force of this challenge, let us return to Freud's description of a repressed unconscious emotion (Freud 1915). In this example, a boy Hans consciously fears his father, and the boy's concern is to harm his father. Hans eventually *displaces* the object of fear (his father) with another object, a horse. He now fears horses, and he is concerned with seeing them harmed or doing harm to them. We can now ask how a concern account of empathy would deal with such a case. Does a concern account of empathy require that an agent match the valences of Hans's unconscious emotions? Would accurate empathy only occur when an agent becomes aware of Hans's *unconscious* concern with harming his father, or would it suffice for the agent to become aware of Hans's *conscious* concern with harming horses? We can reformulate matching relation C on a concern account that requires an *unconscious emotional valence match* between an agent and a target as follows:

> **Matching relation C_2:** An agent's unconscious emotional valences match the target's unconscious emotional valences.

Accurate empathy on concern accounts that employs matching relation C_2 would require that an agent (1) have unconscious emotions whose valences match those of the target; that the agent (2) be aware of the target's unconscious concerns; and that the agent (3) feel motivated to help the target with those unconscious concerns. So when an agent is empathizing with Hans, that agent would, in the first stage, have an unconscious emotional state (in this case fear). The agent would also be unconsciously concerned with harming horses. In the second stage, the agent would then become consciously aware that their own unconscious concern for harming horses is a displaced unconscious concern for harming their father. At this point, the agent may be said to be accurately empathizing

with Hans. But this is a strange result indeed. The agent, presumably, was not aware of their concern with harming their own father prior to empathizing with Hans. The agent may get along perfectly well with his father; or their father may no longer even be alive. The agent may very soon come to believe that these are not in fact his concerns but Hans'. But on an unconscious concern account that espouses matching relation C_2, the agent must at some point instantiate the unconscious emotional states with the same valence as the target and the target's unconscious concerns.

Taking into consideration role of unconscious emotions on a concern account imposes very strong criteria for accurate empathy. Indeed, it would seem like too high a theoretical price to pay for the researcher espousing a concern account of empathy, because these criteria would make it nearly impossible for accurate empathy to occur in everyday situations. For accurate empathy to occur frequently, an agent would require specialized training in order to themselves have unconscious emotions with the same valences as those of a target's unconscious emotions. Further, the agent would have to become aware of the target's unconscious concerns, and feel motivated to help with those concerns.

Here, as in the third reply to the first challenge from unconscious emotions, the researcher espousing a concern account of empathy may reject this second challenge from unconscious concerns and accept the challenge as presented without modifying their account of empathy. But what made the *rejection reply* more appealing on a matching account of empathy was that the agent merely had to represent that content of the target's unconscious emotions. The agent did not have to represent the valences of those emotions. However, on a concern account, an agent would have to themselves have unconscious emotions with valences that match those of the target's unconscious emotions. The agent would also have to become aware of the target's unconscious concerns. These criteria for accurate empathy are different from those of matching accounts. On a concern account then, the implications that follow from including unconscious emotions in one's theory of accurate empathy are more severe. Specifically, if we reject the challenge from unconscious concerns without modifying the criteria for

accurate empathy on a concern account, then we must accept that accurate empathy will almost never occur.

The replies of *eliminativism* and *restrictivism* are also available in reply to the challenge from unconscious concerns. But because of the third criterion of empathic accuracy on concern accounts–that agents feel motivated to help targets with their concerns–it is premature to restrict accurate empathy to conscious concerns at this stage of responding to the challenge. If a target's unconscious emotions cause unconscious concerns, then it is incumbent upon an agent (on a concern account) to be aware of those concerns in order to have the relevant motivations to help that target. In the remainder of the paper, I will present a reply to the challenge from unconscious concerns drawing from Louis Charland's *indeterminacy thesis of affect valence* [ITAV]. By adopting the ITAV, we give up some of the features that unconscious emotions might have. But what we gain is a concern account of empathy that integrates the role of unconscious emotions and makes it possible for accurate empathy to occur much more frequently.

3.6.1 The indeterminacy of affect valence reply

Proponents of Freudian psychoanalysis believe that unconscious emotions significantly shape unconscious motivations and overt behaviour (Suppes and Warren 1975; Rorty 2000).[37] In the previous section we examined a case in psychiatry that exemplified this. In evolutionary psychology, *conscious emotions* are also the cause of *unconscious concerns* (Cosmides & Tooby 2000, p 119, 130; Buss 2000, p 400; Pinker 1997). For example, Cosmides and Tooby (2000) state that the experience of fear causes changes in behaviour that are motivated by concerns that need not be conscious. The example they provide is of an agent walking alone at night and experiencing fear because they are possibly being stalked or about to be ambushed (Cosmides and Tooby 2000, p 3). In this case, the agent consciously experiences fear and is consciously concerned about being stalked or ambushed. But with this conscious experience of fear come many unconscious

[37] Technically Freud does not think that unconscious emotions are not felt. They are what he calls part of the conscious unconscious. The reasons why they remain conscious is a matter of interpretation. But it is important to remember that these emotions are unconscious to the extent that they are *never* noticed or reportable.

concerns such as: a concern for safety; a concern for the location of loved ones; a concern for the location of others that can protect me; a concern for finding a defensive position (*Ibid.*). As in examples of unconscious concerns from psychoanalysis and embodied cognition, this evolutionary psychological example makes salient the force of the challenge from unconscious concerns. Here again the question is whether it is sufficient that an agent become aware of the target's conscious concern about being stalked, or whether an agent must also become aware of the target's unconscious concerns such as the location of others, and finding a safe location. According to evolutionary psychologists, although agents experiencing fear are only consciously concerned with being stalked or ambushed, they also have many unconscious concerns that are functionally and behaviourally equivalent to conscious concerns. And as in the psychoanalytical tradition, the unconscious concerns posited in evolutionary psychology have the same motivational influence on behaviour as conscious ones.

The view that emotions (both conscious and unconscious) can cause unconscious motivational concerns is one that is endorsed by Prinz (2004). I will only briefly summarize his view here, but it is notable because Prinz believes that certain emotions are *intrinsically valenced* (Prinz 2004, p 164). For example, he believes that fear is always negatively valenced. This means that the object of fear is always experienced unpleasantly or as something to withdraw from (Prinz 2004, p 167-168). Prinz believes that fear is intrinsically valenced for evolutionary reasons (Prinz 2004, p 164). Were fear and other "trademark negative emotions" not intrinsically valenced, he believes that "[w]e would gain little [evolutionary] benefit" (*Ibid.*). Fear he says, is an "embodied appraisal" that alerts us to danger (which is always negative). What is of special relevance for present purposes is that Prinz believes that intrinsically valenced unconscious emotions, such as unconscious fear, also exist (Prinz 2004, p 168, 198). On his view, unconscious fear then is functionally and behaviourally identical to fear except that it is not felt or reportable while it occurs.

Prinz cites an experiment on female fear of spiders to support the claim that unconscious emotions with intrinsic valences exist (Prinz 2004, p 203). In this experiment, adult females were asked to perform a series of actions that progressively required closer

interaction with a spider (Arntz 1993). They were first asked to walk towards a spider that was captive in a jar. Then they were asked to touch the jar, to open it, to use a pencil to touch the spider, and so on. Participants could refuse to perform any subtask at any point. Before the tasks, some participants were given an opioid antagonist that blocks brain receptors for endorphin and encephalin that correlated with analgesia and feelings of well-being. The other participants were given a placebo. Unsurprisingly, the participants who were given the drug stopped participating in the experiments prior to those who were given the placebo. However, a surprising result was that both groups of participants reported having subjective experiences of fear that were about the same in qualitative feel. Prinz argues that this difference in the behaviour of the two groups (opioid antagonist compared to placebo) suggests a difference in emotional states, despite conscious experiences of the participants being the same. Prinz thinks that the group which took the opioid antagonist stopped participating in the experiment earlier because they were more strongly motivated by states of unconscious fear. The group that took the placebo continued participating longer. Presumably this was because their fear (both conscious and unconscious) was being influenced by the activity of their unfettered neuroreceptors. Prinz takes this evidence to support the conclusion that intrinsically valenced unconscious emotions influence motivation and behaviour.

As mentioned, the implications of Prinz's view (and others like it) for concern accounts of empathy are significant. If a potential target of empathy has unconscious and intrinsically valenced emotions that cause their unconscious concerns, then for accurate empathy to occur, an agent will not only have to match the valences of these unconscious emotions, they will also have to become aware of the targets' unconscious concerns. The problem here is that accurate empathy would be so infrequent as to render concern accounts of empathy trivial—they would scarcely be explaining empathy at all. A possible reply to this challenge is to sever the connection between unconscious emotions and motivational concerns. To do so, a proponent of a concern account of empathy can appeal to Charland's *indeterminacy thesis of affect valence* [ITAV].

Charland (2005a) investigates what we have already seen to be a central feature of emotional experience: *affect valence*. *Affects* are simply individual felt emotional states.

And we have seen that such states can be classified as either positive or negative. For example, when I experience sadness at the loss of loved one, I *likely* classify this experience as negative. I stress the likeliness of my classification in this example because what Charland argues, contra Prinz, is that the valences of emotional states are not *intrinsic* to those states. They are not intrinsically pleasant and desirable or unpleasant and to be avoided. In other words, the emotional experiences that a target has cannot be described as positive or negative a priori. Emotional experiences are not objectively either positive or negative. Whether a target's emotional experience is positive or negative will depend on the target *attending* to and reporting on their emotional experience. The valence of a particular emotional state (an affect) is "created and structured" by features of second-order awareness of that state (Charland 2005a, p 235).[38] This awareness of an emotional state does not create the phenomenology of the state. Rather, it shapes what an emotional state means to us, and hence shapes its valence. It follows that an agent that is attempting to empathize with a target cannot be accurately aware of the target's emotion state valence prior to the target attending to their emotional experience. This is because, prior to the target attending to their experience, their emotional state valences are *indeterminate*.

It is important to emphasize that the main point of the ITAV—that valence is not an intrinsic property of affects—is opposed to Prinz's view mentioned above. On Prinz's view, some emotions, like fear, are always intrinsically negatively valenced. The negative valence of fear is always a property of fearful states. On the other hand, the ITAV makes it possible that some fearful states are not intrinsically valenced. For example, a target that is exhilarated by fear while skydiving may classify the valence of this affect as positive. On Charland's view, it is the interaction between attention and phenomenal experience that fixes the valence of an affect as either positive or negative. Valence is *indeterminate* prior to attention and report. It is a dynamically created property.

[38] It is important to note that valence is created and structured by features of second-order awareness. The content of the valence—whether a particular emotional state will be valenced positively or negatively—will be dependent upon other factors including development and learning throughout the history of the agent.

Returning to the challenge: does accurate empathy (on concern accounts) require that an agent recognize both the apparent conscious concerns and the less apparent unconscious concerns of a target? The researcher espousing a concern account of empathy who also espouses the ITAV can reply "no" because unconscious emotions are non-valenced, and hence non-motivational states. Unconscious emotional states are states that occur independent of conscious attention or awareness. It follows that if the valence of a *conscious* emotional state is indeterminate prior to awareness, and an unconscious emotional state is one that occurs independent of awareness, then the valence of an *unconscious* emotional state is also indeterminate. We can now see that the implications of the ITAV for replying to the challenge from unconscious concerns are significant. It is compatible with the ITAV that unconscious emotional states exist. But it is not compatible that unconscious emotions are motivational–they do not cause *unconscious concerns*. By giving up on the claim that unconscious emotional states can have the "same informational function" as conscious emotional states, a theorist can reply to this second challenge. Accurate empathy does not require an agent to have unconscious emotional states whose valences match those of a target because unconscious emotional states are not yet valenced. Furthermore, because unconscious emotional states do no cause unconscious concerns, an agent does not need to become aware of a target's unconscious concerns caused by unconscious emotions.

3.7 Concluding remarks

We have seen that a theorist espousing a matching account of empathy can choose from at least three replies to the first *unconscious matching challenge*. First there was *eliminativism*, according to which unconscious emotions do not exist, and thus a matching relation between an agent and target's unconscious emotions cannot be established. Second we considered *restrictivism,* according to which an agent does not need to match a target's unconscious emotional states in experience because unconscious emotional states do not have experiential content. And third we considered rejecting the challenge without modification to the matching account. According to this last reply, accurate empathy *would* require that an agent represent a target's unconscious emotions. But we also saw that this third reply faces two difficulties. The nature of an agent's

access to these representations is not sufficiently described. And the requirements of specialized knowledge and training that this reply imposes would make the occurrence of accurate empathy infrequent.

Next I argued that by incorporating Charland's indeterminacy thesis of affect valence, a theorist espousing a concern-account of empathy can reply to the *unconscious concern challenge*. This allowed us to see that unconscious emotions are not functionally and behaviourally equivalent to conscious emotions. Unconscious emotions are not *motivational* because they are not valenced, hence they do not cause *unconscious concerns*. That a theorist can meet these challenges from unconscious emotions is good news for empathy research. It means that both matching accounts and concern accounts of empathy can take into consideration the role of unconscious emotions in empathy and be equipped with a good theory of empathic accuracy. My care account, developed in the previous paper, benefits from this insofar as it can meet the unconscious concern challenge by adopting the indeterminacy of affect valence reply.

References

Adams, Frederick. “The Informational Turn in Philosophy.” *Minds and Machines* 13.4 (2003): 471–501. Print.

Albert, Stanford CA Gordon H. Bower, others. *Cognition and Emotion*. Oxford University Press, 2000.

Arntz, Arnoud. “Endorphins Stimulate Approach Behaviour, but Do Not Reduce Subjective Fear. A Pilot Study.” *Behaviour research and therapy* 31.4 (1993): 403–405. Print.

Bargh, John A et al. “The Automated Will: Nonconscious Activation and Pursuit of Behavioral Goals.” *Journal of Personality and Social Psychology* 81.6 (2001): 1014–1027. *CrossRef*. Web.

Bargh, John A., and Ezequiel Morsella. “The Unconscious Mind.” *Perspectives on psychological science* 3.1 (2008): 73–79. Print.

---. “Unconscious Behavioral Guidance Systems.” *Then a miracle occurs: Focusing on behavior in social psychological theory and research* (2010): 89–118. Print.

Bargh, John A., and Paula Pietromonaco. “Automatic Information Processing and Social Perception: The Influence of Trait Information Presented outside of Conscious Awareness on Impression Formation.” *Journal of Personality and Social Psychology* 43.3 (1982): 437. Print.

Batson, C. Daniel. *The Altruism Question*. Hillsdale, NJ: Erlbaum, 1991. Print.

Berridge, Kent, and Piotr Winkielman. “What Is an Unconscious emotion? (The Case for Unconscious‘ Liking’).” *Cognition & Emotion* 17.2 (2003): 181–211. Print.

Breuer, Joseph, and Sigmund Freud. *Studies on Hysteria*. Basic Books, 1895.

Charland, Louis C. “Emotion Experience and the Indeterminacy of Valence.” *Emotion and consciousness* (2005): 231–254. Guilford Press, NY: New York. Print.

---. “The Heat of Emotion: Valence and the Demarcation Problem.” *Journal of consciousness studies* 12.8-10 (2005): 82–102. Print.

Cosmides, Leda, and John Tooby. “Evolutionary Psychology and the Emotions.” *Handbook of emotions* (2000): 91–115. Guilford Press, NY: New York. Print.

De Vignemont, Frederique, and Tania Singer. “The Empathic Brain: How, When and Why?” *Trends in Cognitive Sciences* 10.10 (2006): 435–441. *CrossRef*. Web.

Dimberg, U., M. Thunberg, and K. Elmehed. “Unconscious Facial Reactions to

Emotional Facial Expressions." *Psychological Science* 11.1 (2000): 86–89. *CrossRef*. Web.

Doherty, R. William. "The Emotional Contagion Scale: A Measure of Individual Differences." *Journal of nonverbal Behavior* 21.2 (1997): 131–154. Print.

Ekman, Paul. *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*. New York: Times Books, 2003. Print.

Erdelyi, Matthew Hugh. *Psychoanalysis: Freud's Cognitive Psychology.* WH Freeman/Times Books/Henry Holt & Co, 1985.

Evans, Jonathan St. B. T. "Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition." *Annual Review of Psychology* 59.1 (2008): 255–278. *CrossRef*. Web.

Evans, Jonathan St.B.T. "In Two Minds: Dual-Process Accounts of Reasoning." *Trends in Cognitive Sciences* 7.10 (2003): 454–459. *CrossRef*. Web.

Freud, S. *Repression. Standard Edition, 14: 141-158*. London: Hogarth Press, 1915. Print.

Freud, Sigmund. *Inhibitions, Symptoms and Anxiety*. WW Norton & Company, 1926. Print.

Gallese, Vittorio, and Alvin Goldman. "Mirror Neurons and the Simulation Theory of Mind-Reading." *Trends in cognitive sciences* 2.12 (1998): 493–501. Print.

Gelder, B De. "Nonconscious Emotions" in Barrett et Al. *Handbook of emotions* (2010). Gilford Press. Print.

Goldman, Alvin. "Two Routes to Empathy." *Empathy: Philosophical and psychological perspectives* (2011): 31. Print.

Goldman, Alvin I. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press, 2006.

Hoffman, Martin L. "Empathy, Justice, and the Law." *Empathy: Philosophical and Psychological Perspectives* (2011): 230–54. Print.

Iacoboni, Marco. *Mirroring people: The new science of how we connect with others*. Macmillan, 2009.

Ickes, William. "Empathic Accuracy." *Journal of personality* 61.4 (1993): 587–610. Print.

---. "Naturalistic Social Cognition: Empathic Accuracy in Mixed-Sex Dyads." *Journal of Personality and Social Psychology* 59.4 (1990): 730. Print.

Ickes, William, and William Tooke. “The Observational Method: Studying the Interaction of Minds and Bodies.” (1988).

Izard, Carroll E. “Emotion Theory and Research: Highlights, Unanswered Questions, and Emerging Issues.” *Annual Review of Psychology* 60.1 (2009): 1–25. *CrossRef*. Web.

James, William. “On Some Omissions of Introspective Psychology.” *Mind* 33 (1884): 1–26. Print.

Kihlstrom, John F. “Conscious versus Unconscious Cognition.” *The nature of cognition* (1999): 173–203. Print.

---. “The Rediscovery of the Unconscious.” *Santa Fe Institute Studies in the sciences of complexity-Proceedings Volume-*. Vol. 22. Addison-Wesley Publishing Co, 1995. 123–123.

Kihlstrom, John F., Terrence M. Barnhardt, and Douglas J. Tataryn. “The Psychological Unconscious: Found, Lost, and Regained.” (1992).

Knafo, Ariel et al. “The Developmental Origins of a Disposition toward Empathy: Genetic and Environmental Contributions.” *Emotion* 8.6 (2008): 737. Print.

Krebs, Dennis. “Empathy and Altruism.” *Journal of personality and social psychology* 32.6 (1975): 1134. Print.

Lazarus, Richard S. *Emotion and Adaptation*. New York: Oxford University Press, 1991.

Maslow, Abraham Harold, Robert Frager, and Ruth Cox. *Motivation and personality*. Eds. James Fadiman, and Cynthia McReynolds. Vol. 2. New York: Harper & Row, 1970.

McClelland, David C., and David A. Pilon. “Sources of Adult Motives in Patterns of Parent Behavior in Early Childhood.” *Journal of Personality and Social Psychology* 44.3 (1983): 564. Print.

Neumann, Roland, and Fritz Strack. “‘Mood Contagion’: The Automatic Transfer of Mood between Persons.” *Journal of personality and social psychology* 79.2 (2000): 211. Print.

Pettit, Gordon. “Are We Rarely Free? A Response to Restrictivism.” *Philosophical studies* 107.3 (2002): 219–237. Print.

Prinz, Jesse. "Is empathy necessary for morality?" *Empathy: Philosophical and psychological perspectives* (2011): 211-229. Prinz, Jess. “The Sensory Bias of Cognitive Phenomenology.” *Cognitive Phenomenology*. Ed. Tim Bayne and Michelle Montague. Oxford University Press, 2011.

Prinz, Jesse. “Against Empathy.” *The Southern Journal of Philosophy* 49 (2011): 214–233. *CrossRef*. Web.

Prinz, Jesse J. *Gut Reactions a Perceptual Theory of Emotion*. Oxford; New York: Oxford University Press, 2004. *Open WorldCat*. Web. 17 Dec. 2014.

Ramsey, William, Stephen Stich, and Joseph Garan. “Connectionism, Eliminativism, and the Future of Folk Psychology.” *Philosophy, Mind, and Cognitive Inquiry*. Springer, 1990. 117–144.

Rorty, Amélie Oskenberg. “Freud on Unconscious Affects, Mourning and the Erotic Mind.” *The Analytic Freud, Philosophy and Psychoanalysis*. Ed. Levine P. Miachael. London; New York: Routledge, 2000.

Rosenthal, David M. “The Independence of Consciousness and Sensory Quality.” *Philosophical Issues* 1 (1991): 15. *CrossRef*. Web.

Sagi, Abraham, and Martin L. Hoffman. “Empathic Distress in the Newborn.” *Developmental Psychology* 12.2 (1976): 175. Print.

Semin, G. R., and Eliot R. Smith. *Embodied Grounding Social, Cognitive, Affective, and Neuroscientific Approaches*. Cambridge; New York: Cambridge University Press, 2008. *Open WorldCat*. Web. 17 Dec. 2014.

Simner, Marvin L. “Newborn’s Response to the Cry of Another Infant.” *Developmental Psychology* 5.1 (1971): 136. Print.

Sober, Elliott, and David Sloan. Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, Mass.: Harvard University Press, 1998. Print.

Suppes, Patrick, and Hermine Warren. “On the Generation and Classification of Defense Mechanisms.” *The International Journal of Psychoanalysis* (1975).

Van Inwagen, Peter. “When Is the Will Free?” *Philosophical Perspectives* 3 (1989): 399. *CrossRef*. Web.

Westen, Drew. “Implications of Developments in Cognitive Neuroscience for Psychoanalytic Psychotherapy.” *Harvard review of psychiatry* 10.6 (2002): 369–373. Print.

Wilson, Timothy. “Knowing When to Ask: Introspection and the Adaptive Unconscious.” *Journal of Consciousness Studies* 10.9-10 (2003): 131–140. Print.

Wilson, Timothy D. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, Mass.: Belknap Press of Harvard University Press, 2002. Print.

Winkielman, P. “Unconscious Affective Reactions to Masked Happy Versus Angry Faces Influence Consumption Behavior and Judgments of Value.” *Personality and Social Psychology Bulletin* 31.1 (2005): 121–135. *CrossRef*. Web.

Winkielman, Piotr, Paula M. Niedenthal, and Lindsay Oberman, others. “The Embodied Emotional Mind.” *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (2008): 263–288. Print.

Woody, J. Melvin, and James Phillips. “Freud’s Project for a Scientific Psychology after 100 Years: The Unconscious Mind in the Era of Cognitive Neuroscience.” *Philosophy, Psychiatry, & Psychology* 2.2 (1995): 123–134. Print.

Zahn-Waxler, Carolyn, JoAnn L. Robinson, and Robert N. Emde. “The Development of Empathy in Twins.” *Developmental psychology* 28.6 (1992): 1038. Print.

4. Towards an enlarged evolutionary psychological explanation of empathy

4.1 An enlarged evolutionary psychology of empathy

Theories about the evolution of empathy have been proposed in many fields. In primatology for example, de Waal argues that empathy is part of human nature by appeal to empathy occurring in populations of non-human primates (de Waal 1996). Another prominent evolutionary explanation is found at the intersection of philosophy and biology. Sober and Wilson argue that natural selection promotes empathy-mediated parental behaviour that is motivated by concern for the well-being of their children (Sober and Wilson 1998, p 326). They expand the scope of this claim about parent behaviour to claim that empathy in general evolved to cause agents to behave altruistically (at least sometimes) (Sober and Wilson 1998).

In this paper, I focus on the evolutionary psychology of empathy. Of the recent accounts of how, why, and when empathy evolved, the most influential have come from evolutionary psychology. While there exist several evolutionary accounts of empathy (such as those of de Waal, and Sober and Wilson mentioned above), standard evolutionary psychological explanations of empathy differ in two important ways. First, they share a well-developed *theoretical framework*; and second, they use the same explanatory *method*. It is these two features of *standard evolutionary psychology* [SEP] that are referred to when describing its "evolutionary meta-theory" (Buss 2008, p 403). And it is this meta-theory (framework and method) that evolutionary psychologists apply to all the phenomena they provide explanations for.

In the last two papers of my dissertation I developed a care account of empathy that integrates three factors: *values*, *motivational orientations*, and *environmental contexts*. This integration has consequences for evolutionary explanations of empathy. But SEP explanations of empathy have not given these factors serious weight. In fact, the

theoretical tenets of SEP are either antithetical to them or make it possible to not acknowledge their importance.

My aim in this paper is to argue that my care account of empathy best fits an enlarged theoretical evolutionary framework, and that it can begin to be evolutionarily explained by using a revised methodology. This is because my care account makes room for variation in terms of values, environmental contexts, and motivations over time to affect the evolution of empathy. On the other hand, SEP specifies that we should posit psychological mechanisms that have remained fixed since the distant past. I deny that the framework of SEP is sufficient for explaining how and why empathy has evolved. Accordingly, I show how it is possible that the functions and the psychological processes of empathy can change over time as a result of organisms regulating each other's values, and as a result of empathy's use in different environmental context at different times throughout evolutionary history. To explain these possible changes, I propose that the theoretical framework of SEP should be enlarged. As I will show, this has significant impact on the practical method of SEP. My theoretical enlargement and methodological revision of SEP (which I call *enlarged evolutionary psychology* [EEP]), does pose significant challenges to some of SEP's deeply held theoretical tenets and methodological foundations, it does not aim to entirely supplant SEP. Both SEP and EEP share many features. And although the care account and SEP accounts of empathy are similar in their conception of what empathy is, the care account best fits EEP's theoretical framework and can be better explained according to EEP's revised method. The implication here being that SEP accounts of empathy are not as empirically well supported as they are claimed to be.

I begin by examining SEP explanations of empathy. I present SEP's account of empathy, but then go on to show it to be lacking in historical complexity and empirical support. The first step in doing this is to show how SEP explanations of empathy accord with SEP's evolutionary theoretical framework. I focus on three theoretical tenets of SEP:

1) Inclusive fitness

2) Computational theory of mind

3) Slow evolution

I describe these three tenets and how they serve as a mutually supportive *theoretical framework* for SEP. Then, I show how these three tenets have direct implications for SEP's *explanatory method*.

At the end of the paper, I sketch an evolutionary explanation of empathy that is consistent with the account of empathy that I developed in the previous two papers. Attaining this goal will require attaining two sub-goals. First, I will enlarge the evolutionary theoretical framework of SEP to be more consistent with recent evidence from evolutionary biology and computer science. This evidence makes it possible to see how the factors emphasized by my care account of empathy can be explained from an evolutionary perspective. On the other hand, SEP is often not supported by, and more importantly, is often directly at odds with much of this evidence. I will argue that this lack of empirical support justifies a theoretical enlargement of SEP's three tenets to include recent evidence and theoretical advances in biology, computer science, and philosophy.

The second sub-goal is to revise SEP's explanatory method. This method is shaped and constrained by the three tenets of SEP mentioned above. Accordingly, it is characterized by the generation of hypotheses that apply universally; the identification of innately fixed psychological mechanisms; and the attribution of biological function to those mechanisms that were selected exclusively during a specific historical epoch—namely the Pleistocene. I will show that the enlarged evolutionary theory of EEP entails a revision of SEP's explanatory method. This revised method is characterized by the generation of hypotheses that apply universally or only locally; the added identification of psychological mechanisms that are modifiable throughout development; and the attribution of biological function to those mechanisms that may have evolved and been selected during a wider range of time.

4.2 Standard evolutionary psychology [SEP] of empathy

In the previous papers of my dissertation I have argued in favour of a *care account of empathy*. If we imagine an agent walking down a street and seeing a target fall off a bicycle, then we can imagine that the agent empathizes accurately with the target when the agent experiences an emotional state with a matched valence (possibly fear), becomes aware of the target's concern (e.g. for medical assistance), and is motivated to help the target with that concern. On my account, it is the agent's *care* for a target that explains the agent's motivation to help the target with their concerns. If the agent did not care for the target, the agent might still help the target with their concerns. For example, an agent may help the target who just fell off their bicycle because the agent believes it is their duty to do so. But it is the agent's minimal care for the target that often accounts for why the agent and the target share an emotional state with the same valence, and for why the agent is motivated to help the target when empathy occurs. My account places a great deal of emphasis on the role of values, motivational orientations (e.g. cooperation, competition), and environmental context (e.g. time factors, settings, and roles). What is significant about this emphasis is that these three factors influence when empathy occurs and inform us about the conditions in which the occurrence of empathy is possible and more likely. Also as described in the previous papers, my *care account* can be considered an instance of a *concern account* type of empathy (Batson, 1991; Hoffman 1994) rather than the *matching account* type (Ickes 1993; Prinz 2011). It is an account that involves an agent becoming aware of a target's concerns; and crucially, it also involves the agent feeling motivated to help that target with their concerns.

The task of making the care account consistent with evolutionary theory is made easier by my use of the term 'empathy' in that I use it to refer to the same phenomenon as that referred to by many prominent evolutionary psychologists. For example, Simon Baron-Cohen (2003) states that:

> Empathizing is the drive to identify another person's emotions and thoughts, and to respond to them with an appropriate emotion. Empathizing does not entail just the cold calculation of what someone else thinks and feels (or what is sometimes called mind reading). Psychopaths can do that much. Empathizing occurs when we feel an appropriate emotional reaction, an emotion *triggered by* the other person's emotion, and it is done in order to understand another person, to predict their behavior, and to connect or resonate with them emotionally.
>
> Imagine if you could recognize that Jane is in pain but this left you cold, or detached, or happy, or preoccupied. That would not be empathizing. Now imagine you not only see Jane's pain, but you also automatically feel concern, wince, and feel a desire to run across and help alleviate her pain. This is empathizing. (Baron-Cohen 2003, p 28)

Baron-Cohen uses 'empathy' to refer to the phenomenon whereby an agent identifies a target's concerns, experiences a similar or "appropriate" emotion, and feels motivated to help the target with their concerns. It is not enough for the agent to match the target's thoughts or feelings, the agent must also feel motivated to help the target. Baron-Cohen's account, like mine, is a concern account.

Steven Pinker, another prominent evolutionary psychologist, also uses empathy to refer to much the same phenomenon. When addressing the relationship between empathy and helping behaviour, he describes empathy as inducing a state of sympathy that has a strong motivational component:

> …adopting someone's viewpoint, whether by imagining oneself in his or her shoes or imagining what it is like to be that person, induces a state of sympathy for the person (which would then impel the perspective-taker to act altruistically toward the target if the sympathy-altruism hypothesis is true as well). (Pinker 2011, p 541)

Although Baron-Cohen and Pinker's use of the term empathy is not shared by all evolutionary psychologists, they have done the most research on the topic within this field; and they both share my commitments that care for the well-being of a target and a motivation to help are important to empathy.[39] Going forward then, the possibility of

[39] For example, Nesse and Lloyd (1992, p 611) state that empathy is "the ability to experience the feelings of other people as if they were one's own". Accordingly, their usage of the term is closer to a matching account. On the other hand, Buss (2008, p 404) is explicitly follows Baron-Cohen usage and states that: "*Empathizing* allows a person to both predict and to care about how others feel. Without empathizing,

conceptual confusion between what I take to be the dominant view on the evolutionary psychology of empathy and my own view can be ruled out based on these shared commitments. Having fixed a common reference on the phenomenon of empathy, I will now present its standard evolutionary psychological explanations.

An SEP explanation of a particular phenomenon seeks to provide an answer to the question of why that particular phenomenon is occurring as an expression of its "evolved architecture" or "evolved psychology" (Barkow et al. 1992). An often repeated motto of SEP is that "our modern skulls house a stone age mind" (Cosmides and Tooby, 1997). What this means is that the evolved architecture of our mind is adapted to an environment—a Stone Age environment—that is different from our contemporary environment. More specifically, SEP holds that the species-typical (or universal) processes of human psychology are adapted to the physical and social environment of the Pleistocene (Barkow et al. 1992, p 5). Evolutionary psychologists believe that it was during this long epoch—between approximately 2.6 million to 12 thousand years ago—that our innate psychological processes became more complex. Human minds very slowly changed to enable us to survive and reproduce in that specific past environment (*Ibid.*).[40] However, since then, our minds have essentially remained fixed.

In the environment of the Pleistocene, humans are said to have faced many recurring *adaptive problems*. Examples of such problems include: detecting predators; eating nutritious foods; finding a suitable place to live; finding and attracting the most appropriate mate. Solving these problems is said to have been important to survival and reproduction. And evolutionary psychologists hypothesize that a set of dedicated psychological processes evolved to solve each of these recurring adaptive problems. These problems have been divided into four types (Buss 2008, p 66):

---

understanding the beliefs and desires of others enables a person to read facial expressions and writhing body movements to understand that 'I can see that you are in pain.' Empathizing allows a person to express the notion that 'I am upset that you are in pain.'"

[40] The technical term for this past environment is *environment of evolutionary adaptedness (EEA)* (Tooby and Cosmides 1992, p 69).

1) Problems of survival and growth

2) Problems of mating

3) Problems of parenting

4) Problems of aiding genetic relatives

SEP aims to show that the *function* of each of our psychological processes is to solve an adaptive problem.[41] To take a just mentioned example, that contemporary humans have a purportedly universal preference for sweet food is caused by an equally universal mental process underpinning the sensation of sweetness whose adaptive function is to cause us to seek, detect, and choose nutritious fruit in the environment of the Pleistocene (Symons 1992, p 138-139; Pinker 1997, p 525; Cosmides and Tooby 1997, p 13). This functional explanation is justified by its appeal to contemporary statistical regularity and by its historical plausibility. In this case, the historically plausible claim is that this mental process was selected because in the environment of the Pleistocene, nutritious fruit was scarce, and its sweetness correlated with a higher level of nutrition (fruits are more nutritious when their sugar content is highest). Later, in sections 3 and 5 of this paper, I will present an analysis of how the evolutionary theory and explanatory method of SEP informs and shapes its functional explanations of phenomena generally. For present purposes, I will turn to presenting SEP's explanations of empathy.

Evolutionary psychologists argue that there exist essential psychological sex differences that explain why females are better at empathizing than males. The most prominent evidence of this results from the experiments of Baron-Cohen and his colleagues (Baron-Cohen 2003; Baron-Cohen et al. 2005; Baron-Cohen 2007; Baron-Cohen 2011). They have developed a test for what they call the "Empathy Quotient" (EQ). The empathy quotient construct in this test is supposed to measure individual differences in how frequently people attempt to empathize and how accurately they do so. This test consists of a 60 item self-report questionnaire. Forty of the questions are empathy related

[41] "The *function* of an adaptation refers to the adaptive problem it evolved to solve, that is, precisely *how* it contributes to survival or reproduction." (Buss 2008, p 40)

("empathy items"), twenty are not ("filler/control items"). Participants in these experiments are mailed the questionnaire and asked to respond to it without spending too much time on each question (Baron-Cohen and Wheelright 2004; Lawrence et al. 2004). Below are few self-report questions on the EQ test:

> 6. I really enjoy caring for other people.
>
> 10. People often tell me that I went too far in driving my point home in a discussion.
>
> 18. When I was a child, I enjoyed cutting up worms to see what would happen.
>
> 31. I enjoy being the center of attention at any social gathering.
>
> 52. I can tune into how someone else feels rapidly and intuitively.

For each question, participants are asked to respond by circling one of four answers: "strongly agree", "slightly agree", "slightly disagree", and "strongly disagree" (*Ibid.*). Answers are divided into two sets and each answer is assigned a numerical value of either 2 or 1 and summed. Participants are then assigned an *empathy quotient* score based on their total. According to these empathy quotient experiments woman are much more empathic than men. Specifically, three times as many women as men score in the higher ranges on EQ tests (Baron-Cohen and Wheelwright 2004; Lawrence et al. 2004).

Baron-Cohen claims that the psychological processes responsible for empathy are better developed in females than in males (Baron-Cohen 2003, p 39). He claims that this is because of the behaviours that males and females performed that contributed to solving adaptive problems during the Pleistocene. He hypothesizes that empathy had four *functions* in that environment of evolutionary adaptation.[42] Namely (Baron-Cohen 2003, p 425-437):

[42] The notion of *function* in use here is one describes the mental processes' (in this case: empathy) causal effect on behaviour. As such it is a *psychological function* that is also *biological* because it is informed by evolutionary theory. But it can also be characterized as *social* function because it describes how society was arranged during the Pleistocene, and the societal effects of the process on that arrangement.

1) "Mothering"

2) "Gossip"

3) "Social Mobility"

4) "Reading Your Partner"

Each of these functions of empathy is related to a different behavioural adaptive problem.

Baron-Cohen claims that a function of empathy was to improve "mothering" by enabling females to become more accurately aware of their children's needs:

> If one considers that good empathizing would have led to better care-giving, then since care-giving can be assumed to have been primarily a female activity until very recent history, those mothers who had better empathy would have succeeded better in 'tuning in' to their infant offspring's pre-verbal emotional and physical needs, which may have led to a higher likelihood of the infant surviving to reproductive age. Hence, good empathy in the mother would have promoted her inclusive fitness (Baron-Cohen 2007, p 218).

The adaptive problem that this "mothering" function of empathy is said to solve is that of becoming aware of and responding to a child's needs (Baron-Cohen 2003, p 430). Baron-Cohen claims that females, as the primary care-givers to children "until very recent history", innately possess a good capacity for empathy which makes them better empathizers than males (on average) (Baron-Cohen 2007, p 218). Female agents empathizing with their children leads to greater *reproductive success* because empathy contributes to the child's survival and quality of life (*Ibid.*).[43] Pinker also believes that that "[women] feel more empathy towards their friends [than men]" (social mobility function); and that the original function of empathy was its "mothering" function (Pinker 2002, p 345; Pinker 2011, p 539).

[43] Baron-Cohen claims that empathy causes a child to feel "securely attached" and that securely attached children learn faster, are more easily accepted by their peers, are more popular, and that they develop more stable relationship throughout their lives (Baron-Cohen 2003, p 430).

The second function that Baron-Cohen attributes to empathy is that of "gossip" (Baron-Cohen 2003, p 433). He claims that because females are better empathizers they are more likely to gossip amongst themselves:

> A second explanation is that females with better empathy might have found it easier to socialize—chat, gossip, network—with other females, thereby being more successful in creating social support for themselves whilst engaged in being a care-giver to their infant. Social support from other females is likely also to buffer mothers from the range of life events (illness, poverty, loss, physical attack, etc.) that might otherwise threaten her ability to care for her offspring, and so increase the likelihood of her infant surviving to reproductive age, thereby increasing her inclusive (Baron-Cohen 2007, p 218).

The "gossip" function of empathy is said to solve the adaptive problem of forming alliances (Baron-Cohen 2003, p 434). In general, alliances are said to contribute to reproductive success by gaining acceptance and support from others in difficult times.[44] Baron-Cohen claims that females form alliances mostly by empathizing with each other. On the other hand, he claims that male social relationships are more influenced by dominance hierarchies (Baron-Cohen 2003, p 142).

The third function of empathy that Baron-Cohen attributes to empathy is that of "social mobility" (Baron-Cohen 2003, p435). Here he claims that empathy helps females build relationships with members of the population that are less genetically related than others (Baron-Cohen, p 435):

[44] "The survival advantages of having good friends is that you have social alliances and help when the going gets tough. A high-empathizing female, engaged in childcare, is better equipped to create a community of friends who could watch over her children when she is unable to keep an eye on them all the time." (Baron-Cohen 2003, p 427)

> Among humans (and other great apes), males tend to stay in their birth group, while females tend to move to their mate's community. Males therefore are surrounded by their kin more often than females are, and of course they know their kin well, and vice versa. So there may have been less pressure on males to develop good empathy if males typically have had to put far less effort into building and maintaining relationships. Making relationships with individuals you are not genetically related to requires much greater sensitivity to reciprocity and equality, since these are relationships that you cannot take for granted. A woman with low empathy might have had a much harder time being accepted by her in-laws, and earning their support (Baron-Cohen 2003, p 435).

The "social mobility" function of empathy is said to solve the adaptive problem of creating good relationships with others. He states that those relations between organisms that are relatively distantly genetically related "cannot be taken for granted" and so require more empathy to create and maintain (*Ibid.*).This accords with his description of behaviour in the Pleistocene environment where females interacted with genetically more distant organisms more often than males.

Of note is a different "social" function of empathy that is posited by Kruger:

> [E]empathy could arise as a consequence of attachment-related cues that signaled relatively high genetic commonality in the environment of our evolutionary adaptation. An action benefiting genes could be altruistic in terms of the cost to the helping individual (Kruger 2001, p 2).

He states that a consequence of empathy is that it promoted helping behaviour among organisms that were relatively more genetically related. In attributing this "social mobility" function to empathy, Kruger cites Rushton's *genetic similarity theory* (Rushton et al. 1984; Rushton 1989; 1991). On this theory, agents give preferential treatment to targets that are more genetically similar to themselves because they are motivated by innate phenotypic matching mechanisms and feature detectors (Rushton 1989, p 505). On this view, it follows that an agent is more likely to empathize with target A rather than target B if target A has more genes in common with the agent than target B (Rushton (1991).

The fourth and final function that Baron-Cohen attributes to empathy is "reading your partner (Baron-Cohen 2003, p 436). His claim here is that females were

responsible for avoiding aggression and maintaining a good relationship with their mate:

> Women who had a talent for decoding their male partner's next move would have had greater success in avoiding spousal aggression. Women who were good at detecting deception would have also been more skilled at finding sincere males to mate with, and at judging whether a man would treat them well or just impregnate them… Being able to empathize with one's partner also makes one more compassionate and tolerant, which can prolong the life of the relationship (Baron-Cohen 2003, p 436).

This function of empathy contributes to solving the adaptive problem of finding suitable mates. Specifically those mates that are less deceptive and less aggressive. According to Baron-Cohen, males were more aggressive than females during the Pleistocene. And the dominant male reproductive strategy is one of impregnating as many females as possible while disregarding their feelings (Baron-Cohen 2003, p 134). Thus, one of the reasons why females today are better empathizers than males is that it enabled them to "read" which partners would be more supportive and less aggressive.

According to hypotheses generated by SEP, empathy was selected for during the Pleistocene because its five functions had psychological, biological, and social effects that increased reproductive success.[45] The consequences of empathy were that it solved adaptive problems such as: (1) raising children; (2) building alliances; (3) interacting with organisms who share fewer genes; (4) benefiting organisms that share more genes; and (5) choosing suitable mates. And SEP claims that four of empathy's functions—"Mothering", "Gossip", "Social Mobility", and "Reading Your Partner"—explain why females are better at empathy than males.

As mentioned at the outset, SEP explanations of empathy are not the only evolutionary explanations of empathy that exist. But they are by far the most influential inside and outside of empathy research. The papers about the evolution of empathy by Baron-

[45] More specifically, Baron-Cohen claims that good empathy promoted *inclusive fitness*. According to SEP, what makes any phenomenon *adaptive* is that it contributes to inclusive fitness. This is an important part of the evolutionary theoretical framework of SEP. I will return to this issue in the next section when I discuss the three theoretical tenets of SEP.

Cohen, Kruger, and Rushton that I cite above are among the most widely cited. The 2011 book in which Pinker expounds his views on empathy is a New York Times best seller and very widely cited. Also, SEP explanations of empathy have attracted widespread media attention. Pinker's views on empathy appear in the The New York Times (Pinker 2012). Baron-Cohen's appear frequently in newspapers such as the New York Times and the Guardian (Baron-Cohen 2005; 2011a; 2011b). But the mainstream influence of SEP is (of course) by no means limited to its explanations of empathy. Standard evolutionary psychological explanations of phenomena including (among many others) economic behaviour (Hordijk 2014), sexual jealousy (Khazan 2014), fear of immigrants (Smith 2014), body image (Russell 2014), and violence (Mohan 2014) appear frequently in major media outlets. And a cursory search reveals that the mainstream influence of SEP explanations of any phenomenon far outweighs that of non-SEP explanations informed by evolutionary theory. I believe that the popularity of SEP explanations is in large part due to its stable evolutionary framework and explanatory method. The theoretical framework of SEP has three main tenets. These tenets constrain the explanatory method of SEP, which together make for a powerful tool allowing researchers to generate novel and consistent explanations of contemporary behaviour.

In the next two section I turn to an examination of the three tenets of SEP and of their relationship to SEP's explanatory method. This will be done with the aim of enlarging the evolutionary framework of SEP and drawing out this enlargement's implications for a revised explanatory method. I call this new approach *enlarged evolutionary psychology*. I will then apply this approach to provide a better evolutionary explanation of empathy.

4.3 The evolutionary framework of SEP: Three tenets

The paradigm of standard evolutionary psychology is informed by three mutually supportive theories:

1) Inclusive fitness

2) Computational theory of mind

3) Slow evolution

Taken together, these three tenets provide a consistent framework which informs all SEP explanations of contemporary behaviour. Tenets (1) and (3) were developed in evolutionary biology, and (2) in cognitive science. In this section, I will describe these three tenets and then show how they mutually support each other.

4.3.1 Inclusive fitness

Inclusive fitness is an evolutionary biological theory which predicts that an agent (or organism) gains *fitness* by behaving in a way that favours the reproductive success of targets who share the most genes relative to other targets (Hamilton 1964). On this theory, the term 'fitness' refers to the chances of organism's genes surviving into the next generation. Furthermore, an agent is predicted to behave in such a way as to maximize "inclusive fitness"—which means that the agent will behave to maximize the chances of survival of copies of those same genes found in other organisms. Inclusive fitness theory predicts which genes causing behaviours will be selected for according to the following formula:

$$R > c / b$$

In this formula, [R] is the fraction of genes in common between an agent and a target; [c] is the energy (or caloric) cost to the agent of a particular behaviour; and [b] is the energy benefit to the agent of that same behaviour (*Ibid)*. This means that genes causing behaviours will be favored by natural selection when the *coefficient of genetic relatedness* [R] will be greater than the fraction of the energy cost divided by its benefit. Here, energy cost is a proxy for fitness in so far as benefiting closer genetic relatives promotes survival to reproductive age and reproduction.

Inclusive fitness has been extended to explain seemingly altruistic behaviour. This extended theory is called *reciprocal altruism* (Trivers 1971). Altruistic behaviour can be defined as a behaviour whose goal is to benefit the target of the behaviour as opposed to the agent. Sometimes altruistic helping behaviour can come at a high energy cost to the agent. For example, when an organism comes to the rescue of another or aids another in combat (*Ibid.*). Reciprocal altruism is an explanation of altruistic behaviour that is consistent with inclusive fitness. It explains altruistic behavior as an evolutionary adaptation that maximizes fitness by causing organisms to behave altruistically when helping behaviour is likely to be reciprocated (*Ibid.*). The theory claims that "selection will discriminate against" targets of altruistic behaviour who do not reciprocate ("cheaters") because not reciprocating will have adverse effects on their life (e.g. reprisal by other organisms) (*Ibid.*). Accordingly, seemingly altruistic behaviours—those that appear to be performed at great cost to the agent and little cost the target—are, in fact, instances of reciprocal altruism.[46]

Inclusive fitness is an entrenched tenet of SEP. Evolutionary psychologists claim that that nearly all of the mental processes that are inherited as a result of evolutionary processes conform to the rule of inclusive fitness:

> The realm of adaptive information-processing problems is not limited to one area of human life, such as sex, violence, or resource acquisition. Instead, it is a dimension crosscutting *all* areas of human life, as weighted by the strange, *nonintuitive metric of their cross-generational statistical effects on direct and kin reproduction* (Cosmides and Tooby 2000, p 6 [my italics]).

Inclusive fitness constrains our innate preferences and motivations, which in turn cause our behaviours. Pinker goes so far as to state: "The ultimate goal that the mind was designed to attain is maximizing the number of copies of the genes that created it." (Pinker 1997, p 43) However, it is important to note that evolutionary psychologists *do not* claim that contemporary humans maximize their inclusive fitness with every contemporary behaviour. Rather, it is that our behaviours are *guided* overall to maximize inclusive fitness because the psychological mechanisms we innately possess motivate us

[46] Inclusive fitness is an important component of Dawkins' theory of the "selfish gene" (Dawkins 1976).

to do so (Buss 1995, p 10). So according to SEP, humans in the contemporary environment are motivated by psychological mechanisms that conform to inclusive fitness; but the behaviours that result from these motivations do not always increase inclusive fitness. This is because the contemporary environment differs from that of the Pleistocene (*Ibid.*). Accordingly, SEP describes humans not as "fitness maximizers", but as "fitness strivers" (Tooby and Cosmides 1990, p 420). We can begin to see how inclusive fitness shapes the explanatory method of SEP. But I will be addressing this issue in the next section. For now, let us turn to the second tenet of SEP.

4.3.2 The mind is like a computer

The second tenet of SEP is that the human mind is like a computer. More specifically, the human mind is identical in principle to a *classical* computational device.[47] I say "in principle" because, as evolutionary psychologists point out, there are many different computers in the world. According to SEP then, human minds operate much like those computers, but the differences are that they are slower and more complex (Pinker 1997, p 40). In these ways, the mind is not like everyday computers. The ways in which human minds are literally like or identical to everyday computers are in the ways that many everyday computers instantiate specific principles of *design* and *operation*. I will address these features in turn.

Two important elements of computer design are *hardware* and *software*. Computer hardware is the set of physical elements that constitute the system. This set includes items like power supplies, hard drives, processing units, and circuitry. According to SEP, the *brain* is the design equivalent of a computer's hardware (Barkow, Cosmides, and Tooby 1992, p 7). Software is set of symbols and rules according to which those symbols are manipulated. These manipulations govern the operation of the hardware. For SEP, the *mind* is the design equivalent of computer's software. In this way, SEP claims that the mind is like a computer. The *computational theory of mind* employed by SEP originated

[47] A classical computational device is one which operates according to Turing computation as opposed to connectionist computation (Fodor and Pylyshyn 1988; Fodor 1989). Some researchers believe that connectionist systems do (or can) operate according to classical computation (Chalmers 1990, 1993; Clark 1993). But no consensus has been reached on this issue (Berkeley 1997).

in disciplines such as computer science, philosophy, and linguistics, among others (Pinker 1997, p 24). Theorists in the disciplines that espouse these views have now coalesced in the field of *cognitive science* (Adams 2003). Accordingly SEP integrates this interdisciplinary approach which holds that the mind is an *information* processing system, and that this system is *functionally* realized in the brain. To say that the mind is an information processing system is to say that it receives information from the environment; that this information is translated into a symbolic language; and that the rule-based manipulation of these symbols causally guides behaviour in virtue of its symbols representing environmental information (*Ibid*). To say that the mind is functionally realized in the brain is to say that the symbolic language or software of the brain is spread out across many regions of the brain, and that it can be multiply realized in various parts of the brain. In sum, the brain is to hardware as the mind is to software.

For SEP, the software of the mind is composed of many programs. It is composed of thousands of separate and specialized computational programs or mechanisms (Barkow, Cosmides, and Tooby 1992, p 39). Each of these programs, also called *modules*, operates in relative isolation from the others by being responsive to only certain environmental information—modules are *domain-specific*. And what made them this way is our genes' past exposure to the specific recurrent adaptive problems of Pleistocene life. Each module's domain of information processing is restricted as a result of our ancestors having solved the same adaptive problems many times. Accordingly, each of the mind's programs is selected for its function of causing adaptive behaviour in the environmental context of the Pleistocene. The properties of modules and their architectural arrangement in the mind are subjects of ongoing debate (Samuels 1998). What is important to emphasize for the present purpose is that, according to this tenet of SEP, the mind is like a computer that runs many programs at the same time, and that these programs are isolated (informationally) from each other. And whereas the programs of everyday computers have been coded by programs and installed by users, SEP claims that the programs of the human mind are coded by our genes and installed by natural selection (Pinker 1997, p 23). Accordingly, the design and operation of our minds is specified by an innate and universal "genetic program" that was shaped by natural selection to cause

specific behaviours—namely, those behaviours that maximized reproductive success in the environment where all humans are either hunters or gatherers (Pinker 1997, p 21). As we shall see, this characterization of the Pleistocene's environment as being populated by hunters and gathers will be relevant to SEP's claims about essential sex differences in empathic capacity. It will turn out that because males were mostly hunters and females were mostly gatherers, their minds were, and continue to be different.

That the mind is like a computer is a central tenet of SEP. Evolutionary psychologists claim that the entire mind operates according to the principles of design and operation of a computer. Pinker goes so far as to state that:

> Without the computational theory, it is impossible to make sense of the evolution of the mind… A program is an intricate recipe of logical and statistical operations directed by comparisons, tests, branches, loops, and subroutines embedded in subroutines… Human thought and behavior, no matter how subtle and flexible, could be the product of a very complicated program, and that program may have been our endowment from natural selection (Pinker 1997, p 27).

According to SEP, the program that is the mind has very specific properties. It is innately specified. Each human is born with a genetic code that causes the mechanisms of the mind to develop into a fixed state. And this fixed state is historically stable. The way the mind's mechanisms operate and are arranged has remained the same ever since their selection during the Pleistocene. In section (5) of this paper, I will argue that this tenet of SEP should be enlarged to include recent developments in computer science. These developments shed new light on the possible operation on the mind, and they allow for new explanations of how it evolved. But first I will turn to the last theoretical tenet of SEP.

4.3.3 Evolution is slow

According to SEP, the strongest process of biological and hence psychological change throughout history is *natural selection*. It is an evolutionary process in virtue of which heritable variations continue to exist. For SEP, these variations are individual genetic differences that are passed on from parents to children during reproduction. Thus, genes are the units upon which natural selection effects its selectivity by changing their

frequency within a species. Other evolutionary processes such as mutation and drift are also responsible for changes in heritable differences, but their effect is minimal. Changes in our genetic code that determine the universally stable state of our psychological development are the result of natural selection. And, according to SEP, evolution by natural selection is extremely slow.

SEP claims that evolutionary change occurs in small increments over millions of years (Buss 2010). It is said to occur primarily by natural selection: a process by which genes are selected based on their contribution to behaviour causing inclusive fitness (the second tenet) and hence reproductive success during a long period of time in the past. But whether a particular behaviour will in fact maximize inclusive fitness in the present is indeterminate. To compensate for this lack of guarantee, natural selection is said to select for genes that produce psychological mechanisms that are sensitive to environmental cues that are probabilistically associated with maximizing fitness in the past (Tooby and Cosmides 1990, p 406). Natural selection does not select for genes based on their performance in the current environment. Rather, it extracts statistical relationships that are unobservable, and that organisms are generally not consciously aware of (Cosmides and Tooby 2000, p 9). Thus, our mental mechanisms are said to "reflect most closely the actual long-term statistical structure of the ancestral world." (*Ibid.*) This long-term statistical structure includes relationships like that between skin color and group identity (Machery and Faucher 2004), skin color and ovulation (Buss 2008, p 154), snakes and their lethality (Barkow, Cosmides, and Tooby 1992, p 72), sweet foods and their nutritional value (Cosmides and Tooby 1997, p 13), and so on. It takes thousands of generations for evolutionary change to occur because natural selection continually selects genes that lead to the development of psychological mechanisms which reflect such statistical relationships which held in the distant past (Pinker 1997, p 42). In essence, natural selection continues to select for genes that reflect the past insofar as it continues to select for genes are that are probabilistically associated with maximizing inclusive fitness in the past; and according to SEP, ninety-nine percent of the human past was spent living in a stable environment with a stable social arrangement of hunters and gatherers

(*Ibid.*). Hence the dictum of SEP: "our modern skulls house a stone age mind." (Cosmides and Tooby, 1997, p 10).

4.4 Mutually supportive tenets

The three tenets of SEP—*(1) inclusive fitness, (2) computational theory of mind, (3) slow evolution*—support each other. Tenet (3) of SEP implies that natural selection is the only process that can build functional organization into organisms (Cosmides and Tooby 2000, p 4). This is supported by tenet (2) which states that the mind is like a computer, in that computers are functionally organized systems. Each program of the mind contributes to its own naturally selected function. Accordingly, the tenet that the mind is shaped primarily by natural selection (3) and the view that the mind is like a computer (2) mutually support each other. That evolution is slow then (3), also supports the view that mind is a certain type of computer. Specifically, it is a type of computer with an innately specified program. If evolution is slow, then the mind will be a like a computer that has a historically and developmentally fixed program. This program will not have changed for millions of years, nor will it change throughout an organism's life.

Figure 4: *Mutually supportive tenets of SEP*

Tenet (1), which holds that all of our mental programs cause us to strive to increase inclusive fitness, provides an explanation of the selective process (natural selection) which makes our minds like this type of computer (2). Inclusive fitness describes the constraints and goals of the mind's program; it describes why it operates the way it does. Additionally, inclusive fitness allows for the operation of the mind to be mathematically formalized (in principle). For example, if we accept that empathy is an important cause of

helping behaviour, and that the program responsible for empathy conforms to the rule of inclusive fitness, then we can, in principle, determine the conditions in which an agent is more likely to empathize with and subsequently help a target. These conditions will involve (A) considerations having to do with how closely problems or goals in the current environment of the organism resemble the problems and goals that organisms' ancestors faced in the distant past environment (e.g. mating, predators, resource availability) and (B) the quantity of genes likely to be shared between the agent and the target relative to other targets.

In this mutually supporting way, SEP provides a unified theoretical framework for explaining human behaviour as a function of programs or mechanisms that were selected to maximize inclusive fitness in the distant past. The theoretical framework of SEP has important implications for its explanatory method. In the next section, I draw these out.

4.5 The method of SEP

SEP is consistent in its use of the same explanatory method. It is a method that involves generating hypotheses and justifying them in way that is consistent with its three theoretical tenets. I describe the explanatory method of SEP as follows:

1) Choose a widespread phenomenon.

2) Posit innate, universal, and historically fixed mechanisms that cause the phenomenon in (1).

3) Generate and test hypotheses about how the mechanisms in (2) functioned to cause behaviour that solved recurring adaptive problems during the Pleistocene.

4) Justify the supposition that the mechanisms in (2) exist by appeal to statistical normality in (1) and historical plausibility in (3).

SEP calls this method *reverse engineering* (Barkow et al. 1992, p 55; Pinker 1997, p 43; Buss 2005, p 25). The goal of this explanatory method is to explain contemporary human

behaviour. The many evolved programs of the human mind underwrite our abilities and motivate our behaviour. But humans do not have direct access to the computational algorithms of these processes. For example, most human beings have a capacity to see. But we do not have access to the innate functional mechanisms that compose the visual system. Similarly, according to SEP, male sexual interest in their wives diminishes significantly in the first years of marriage; and throughout life, males are very easily aroused by possible female sexual partners other than their wives (Pinker 1997, p 471). But males do not have direct access to all of the causes of these behaviours. SEP explains these behaviours by claiming that males possess psychological mechanisms that cause an "insatiable" desire for sexual variety, and that the function of this mechanism is to increase the number of their offspring (Pinker 1997, p 473; Buss 1994, p 76). The goal then of SEP's method is to pull aside the curtain of illusory motivations of contemporary human behaviour to reveal the true evolutionary ones. It claims that our contemporary behaviours and social organizations are, in a very real sense, the consequence of fixed mechanisms.

The mechanisms posited by SEP to explain a current widespread phenomenon are *fixed*, *innate*, and *universal*. This second step of SEP's method is constrained by the third tenet (evolution is slow) and the second tenet (the mind is like a computer) of its theoretical framework. That evolutionary change occurs slowly contributes to the explanation of why it took so long for our psychological mechanisms to be genetically assimilated. Our non-human ancestors lived for two millions in the stable environment of the Pleistocene. It is only recently that humans migrated from the savannah plains of Africa and that we changed from being hunters and gathers into famers (Barkow 1992, p 5). Accordingly, SEP claims that it is unlikely that novel change in our psychology could occur in this relatively short period since the end of the Pleistocene (only 13 000 years). As mentioned, this is because SEP believes that such changes occur almost exclusively by means of natural selection which is taken to be a very slow and gradual process. Changes in our way of thinking occurred slowly during the Pleistocene and were passed from one generation to the next until they became universal. In this way, the tenet that evolution is

slow constrains SEP to explanations of behaviour that are caused by mechanisms that have been fixed since the Pleistocene, innate, and universal.

The second theoretical tenet of SEP—that the mind is like a computer—similarly constrains its method. As described above, this tenet holds that the human mind is like a computer whose code has been programmed and installed by evolutionary processes. But the type of computer that SEP claims to mind is like here is of crucial importance. As mentioned, the mind is like a classical everyday computer that is slower but more complicated. Such computers use fixed software that has been installed by an outside user. And the software on a computer typically remains unchanged unless it is updated by the user. The point I would like to emphasize is that the type of computer SEP claims the mind to be like is one whose code is installed during the manufacturing process and then left unchanged. The human mind's code then is similarly installed at conception and then remains fixed throughout development. It is a type of computer that is *static*. And it is programmed for the environment of the Pleistocene. It follows that the way it processes contemporary environmental input is always by means of these fixed programs which reflect historical statistical relationships (Tooby and Cosmides, 1990). SEP's theory of what type of computer the mind is like (tenet 2) constrains the type of mechanisms that can be posited by its explanatory method. This does not conflict with the supposition that the mind's program is very complex and responsive, with (as Pinker puts it) a great many conditional subroutines activated by specific internal or external conditions.

The third step of SEP explanatory method involves generating hypotheses about how our current psychological mechanisms functioned to solve adaptive problems during the Pleistocene. Here again we see that the view of evolution being very slow informs SEP's claim that our current psychological mechanisms have remained unchanged since the Pleistocene. The goal then is to "reverse engineer" them in order to determine their evolutionary function in the past. Also, we see here the influence of the first tenet (inclusive fitness). If our current psychological mechanisms were selected during the Pleistocene, then they were selected because they were adaptive. And for them to have been adaptive, according to SEP, implies that they conformed to the rule of inclusive fitness. Accordingly, any statistically normal behaviour under investigation will be

explained by appeal to psychological mechanisms that were selected because they motivated organisms to prioritize their own survival and reproduction and the survival of other organisms with whom they shared the most genes. Furthermore, even if these behaviours no longer contribute to increasing inclusive fitness in the present, SEP's adherence to this tenet constrains its explanatory method to mostly posit psychological mechanisms that increased inclusive fitness during the Pleistocene.

The *fourth* step in the explanatory method of SEP is a justificatory one. It justifies its claims in step two about the functions of those mechanisms causing contemporary behaviour in two ways. The first way is by appeal to the tenet which holds that evolution is slow. It is because evolution is slow that it would take a long time for the development of any particular psychological mechanism to be transmitted throughout and selected within the entire population. Furthermore, any psychological mechanism must have been repeatedly selected throughout history for the functions it performed during the Pleistocene because the Pleistocene is the longest epoch in human history. The second way that this fourth step justifies the functions of the mechanisms it posits is by combining the tenet which holds that evolution is slow with the first tenet (inclusive fitness). This allows SEP to make the function under investigation evolutionarily plausible. For a function to be evolutionarily plausible, the criterion to meet is that it would have contributed to inclusive fitness given the social roles, ways of life, and other biotic and abiotic environmental factors posited by SEP's description of the Pleistocene.

We can now more clearly see how the theoretical tenets and the explanatory method of SEP affect its explanation of when, why, and how empathy evolved. *First*, SEP choses a widespread phenomenon: empathy. Many animals care about the well-being of certain other organisms. They become aware of those organisms' concerns, and they feel motivated to help those organisms with their concerns. In the *second* step, SEP then posits that the causes of this phenomenon are psychological mechanisms. These mechanisms are assumed to be innate (developmentally fixed), and universal. They are responsible for causing the way in which organisms are sensitive to their environment, and their activation in the presence of certain inputs will cause empathy. In the third step, SEP generates hypotheses about the adaptive consequences (or functions) of empathy

during the Pleistocene. As we have seen, SEP hypothesizes that the functions of empathy in the past were that it contributed to raising children (mothering), forming alliances (gossip), creating and maintaining relationships with organisms having fewer genes in common than other organisms (social mobility), and avoiding male aggression (reading your partner). The generation of these hypotheses is constrained by SEP's first tenet: inclusive fitness. Thus, each of empathy's functions will have increased inclusive fitness. Experiments are then devised to then test whether the distant historical functions of empathy are still having consequences in the contemporary environment. As described above, Baron-Cohen's experiments involve a set of self-report questions.[48] Within the theoretical framework of SEP, the results of these experiments provide evidence for widespread innate sex differences in the psychological mechanisms causing empathy. These differences in the innate psychological mechanism are justified in step (4) by appeal to their historical plausibility. Here, the tenet that evolution is slow and that the mind is like a computer support the claim of historical plausibility. The distant historical environment remained relatively unchanged for a long time. Thus, the psychological mechanisms that caused behaviour also became fixed over time. The "innate structure of the evolved neural machinery" became more complex little by little (Buss 2005, p 30).

4.6 An enlargement of theory

The outcome of SEP's method is that for any phenomenon it seeks to explain in the contemporary environment, it will posit the existence of innately fixed psychological mechanisms that cause this behaviour. This method allows us to generate explanations for why humans are motivated to behave the way they do. And it is indeed informed by evolutionary theory. But I have aimed to show that the tenets that SEP adheres to, and the way these tenets shape its method, results in a consistent but very a specific type of evolutionary explanation. First, SEP explanations of behaviour are strongly individualistic. The sensitivity to particular features of the environment, the preferences, and the motivations that humans possess all originate from the individual's psychological mechanisms and their arrangement or architecture. Second, these psychological

[48] These self-report questions are generated seemingly independently from any relationship to SEP's tenets or method.

mechanisms motivate behaviours that promoted inclusive fitness in the past. And third, this universal structure of the human mind is developmentally and historically fixed. The structure of the human mind that SEP posits is present in a human's genetic code at the moment conception. This structure causes psychological mechanisms to develop, on average, into a fixed state regardless of changes in the environment and individual experience. The functions that these mechanisms are selected to perform throughout history are also fixed, and they have been shared by humans throughout history. Accordingly, these mechanisms constitute an unchanging and "essential" human nature (Baron-Cohen 2003, p 306).

There exist many important critiques of SEP. The most influential of these is that of Buller (2005). He is critical of all three of SEP theoretical tenets and concludes that SEP "is wrong in almost every detail." (Buller 2005, p 481). Buller challenges the data supporting the view that parental love and care is merely a function of genetic relatedness (Buller 2005, Ch. 7). He is critical of the claim that the mind is composed of modular components that are universally fixed throughout development (Buller 2005, Ch. 4). And he is critical of the claim that, because evolution is slow, the minds of contemporary humans is for the most part entirely the same as that of humans that who lived during the Pleistocene (Buller 2005, p 142). Buller is also critical of SEP's method. Specifically he claims that we cannot know enough about our very distant evolutionary past to enable us to specify precisely the recurring adaptive problems that humans faced (Buller 2005, p 93).

Similarly, Stotz and Griffiths have argued that SEP's claims about the functions of mental mechanisms during the Pleistocene are tenuous (Stotz and Griffiths 2002, p 13). They argue that SEP cannot adequately describe the functions of mental mechanisms in the *environmental niches* of Pleistocene humans without first knowing about the structure their minds (Stotz and Griffiths 2002, p 14). And knowing about this structure requires knowing about the "lifestyle" of humans that lived during that time. Thus, without accurate knowledge of Pleistocene ways of life, there will not be accurate knowledge of the Pleistocene environment (*Ibid.*). And, like Buller, they point out that this evidence may not yet or ever be available.

Feminist evolutionists are also critical of SEP along these lines. For example, Liesen argues that the gender roles posited by SEP's description of the Pleistocene are unjustified (Liesen 2011, p 1). For example, she claims that female preferences for mates with high status and resources are better explained by appeal to *environmental variables* and *social structures* rather than innate psychological mechanisms (Liesen 2007 cf. Buss and Malamuth 1996, p 12).

I have just referenced three important critiques of SEP (Buller 2005; Stotz and Griffiths 2002; Liesen 2007, 2011). But there are many others.[49] They are each critical of some elements of SEP's theoretical framework, and in their own right, address important conceptual ambiguities and epistemological challenges faced by SEP. Some of these criticisms argue that evolutionary psychological explanations should reject or shift emphasis from a gene-centered nativist approach to one that is more sensitive to developmental processes (Greene 2004; Stotz and Griffiths 2002; Griffiths 2007), while others argue that the theoretical framework of SEP should be enlarged in order to include new developments in evolutionary theory such as *niche construction* (Stotz and Griffiths 2002; Sterelny 2003).[50] Throughout however, the impetus is towards the conclusion that SEP is leaving something out. These criticisms suggest (1) a renewed emphasis on context (both cellular and outside the organism), and (2) a new focus on the interactive relations between organisms and their environments at different time scales (Caporeal and Brewer 2000, p 25; Barker et al. 2014). However, while some of these criticisms argue that the theoretical framework of SEP should be enlarged, the theoretical framework of SEP has remained entirely intact (Machery and Berrett 2006; Buss and Schmitt 2011). In other words, outright rejections of SEP have failed to convince its practitioners to abandon the paradigm, and attempts at theoretical enlargement have left the basic tenets of SEP unchanged.

I believe that there are two important reasons for this intransigency in the theory and method of SEP despite its many critics. The first is that the three tenets of SEP form a

[49] For example, see the anthology by Scher and Rauscher (2003).
[50] I will address this development specifically below.

consistent and mutually supportive whole (as shown in section 4.3). Existing attempts at enlargement, though largely correct in my view, do not show how new developments in evolutionary theory either integrate with or replace the existing tenets of SEP in a consistent manner; they do not sufficiently describe the new theoretical relationships that enlargement would entail. Furthermore, for a particular enlargement proposal to be accepted, it would have to present theoretical tenets that were *at least as* mutually supportive as SEP's current tenets. Existing attempts at enlargement do not describe how the elements of its proposed enlargement would provide a mutually supportive meta-theory for SEP. The second reason I believe existing enlargement proposals and criticisms of SEP have failed is that they have not clearly drawn out the implications that enlarging or rejecting elements of SEP's theory have for its explanatory method. As I showed in section 4.4, the three tenets of SEP shape and constrain its explanatory method. Current criticisms of SEP and attempts at theoretical enlargement have not shown how their proposed modifications would impact the explanatory method of SEP in practice.

In the next section I propose a theoretical enlargement of the theoretical framework of SEP. While parts of this proposal have been argued for in existing criticisms of SEP, others are new.[51] I will show how the enlarged theory of SEP that I propose constitutes a consistent and mutually supportive evolutionary meta-theory for an *enlarged evolutionary psychology*. Then, I will show how this enlarged theory has significant implications for evolutionary psychology as it is standardly practiced. In light of these, I will revise the method of SEP in such a way that can be used by practicing evolutionary psychologists. In doing so, I will show how EEP provides better evolutionary explanations of empathy than SEP.

4.6.1 Niche construction and mutualism

The first enlargement to the theoretical framework of SEP that I propose includes *niche construction theory* [NCT]. NCT is a recent development in evolutionary biology that emphasizes the capacity that organisms have to modify natural selection in their

[51] I will indicate these where applicable.

environment and thereby affect their own, and other species' evolution. Niche construction occurs when organisms modify their own and/or each other's environments (*niches)* by their metabolism, their activities and their choices (Odling-Smee et al. 2003, p 419). Rather than treating the environment as a fixed place that recurrently presents a limited set of problems to which organisms passively adapt, NCT treats the environment as a modifiable; organisms construct and modify their environments in order to solve existing problems which in turn creates new problems as an effect (Laland and O'Brien 2010, p 2). An important theoretical insight of NCT is that selection pressures are not only to be found in the environment independent from organismal activity. Causal effects feeding forward from organisms into their environments and back from these modified environments into the same or other organisms create new selection pressures and adaptations. Accordingly, NCT is a selective process, a major cause of evolutionary change.

Examples of niche construction include the many ways in which organisms define and alter their environments. For example, excreting waste products (Laland et al. 2000, p 165); migration or dispersal (Laland and Brown 2006, p 98); fungi manufacturing decomposing organic matter (Odling-Smee et al. 2003); pupal cases (Gullan and Cranston 2009); nests and burrows (Brown 2014). The significance of these examples should not be underestimated. Niche construction is so frequent that it is surprising it took so long for its effects on natural selection to be appreciated. For example, there are 9500 types of ants and 2,000 types of termites that build nests; 20,000 types of solitary bees and many social bees also construct nests; 7000 types of fly are known to construct shelters; 1,800 types of earwigs build nests; 140,000 types of butterflies and moths build pupal cocoons (Laland and Sterelny 2006). And the list of niche constructing insects goes on. But the insects are certainly not the only niche constructers. The vast majority of birds construct nests; fish also construct nests, and spawning sites; lizards, moles, rabbits, and countless other vertebrates actively modify natural selection by modifying their environment (*Ibid.*). As Laland and Sterelny (2006) put it, "[t]he ubiquity and impact of niche construction is no longer open to question." (*Ibid.*)

The importance of niche construction can be seen in its physiological and social impact. Not all of an organism's activities should be considered niche construction. Only those activities that have significant evolutionary consequences should be included. Such activities affect the physiology of organisms as well as the biotic and social environment in which they live. The niche construction activities of one generation of organisms affect an environment which may in turn affect subsequent generations of those organisms or other types of organisms which come upon this constructed environment during the same or later time period (Odling-Smee et al. 2003). For example, the limbs of certain borrowing frogs have become increasingly specialized as a consequence of the presence of burrows left by their ancestors (*Ibid.*). These frogs began constructing small burrows. The burrows were then encountered and further constructed by subsequent generations of frogs. The presence of these burrows created a new selection pressure which resulted in the frogs developing different limb structures that allowed them to burrow more efficiently and create more complex shelters. This is clear example of causal feedback between the organisms modifying their environment and the environment having a modifying effect on the morphology of those organisms. A different example lends credence to the claim that niche construction also has modifying effects on the operation of an organisms' brain and hence its psychology. Certain spiders exhibit specialized behaviour for threat detection and communication which occur only on their webs (*Ibid.*). It is plausible to hypothesize that these behaviours and the psychological processes that make them possible became increasingly specialized as a result of being born with a propensity to create webs in an environment where complex webs were already present.

By modifying their *abiotic* environment, organisms modify the selection pressures they face, which results in evolutionary changes in their biology. But organisms also modify selection pressures by modifying the biotic environment in which they live through *mutualistic interactions*. Organisms are usually born in an environment populated by other organisms of the same species or type. Throughout their lives, they compete and co-operate through shifting alliances and social organizations that affect their survival and reproduction (Barker 2008). For example, many organisms exhibit cooperative behaviours (e.g. pelicans hunting with other pelicans of no close kin relation) (Clutton-

Brock 2009). But in many instances, organisms interact with other organisms of a different type in equally evolutionarily significant ways. Giant Moray eels and Groupers are known to frequently develop *hunting associations* (Bshary et al. 2006). So do Black-backed jackals and honey badgers, and wolves and honey badgers (Clutton-Brock 2009). Furthermore, *foraging associations* among non-kin and organisms of different species are well documented (Dickman 1992). These examples show that organisms modify the organism relations in their environments in order to solve shared adaptive problems. That they do so suggests they possess evolved psychological processes enabling such interspecific communication and flexible coordination evolved as a result organisms modifying their social environment.

What NCT importantly shows for present purposes is that natural selection does not operate only according to inclusive fitness (the first tenets of SEP). Inclusive fitness theory states that genes causing behaviours are selected to contribute to the reproductive success of those genes according to a rule which quantifies over the variables of genetic relatedness and energy costs ($R > c / b$). Selective pressure originates in the environment and is exerted on organisms which adapt according to this rule. Doing so increases the likelihood that their own genes (or other copies of them) will be transmitted to the next generation. However, NCT shows that evolutionary change is not this simple. Selective pressure is not always exerted independently of the activities of organisms. Organisms modify their environment in ways that create causal feedback loops which exert selective pressure on them, other organisms, and subsequent generations of organisms. And, as seen in the case of mutualistic interactions, they shape their social environment in ways that promote survival and reproduction, but that do not appear to conform to inclusive fitness.

Within the theoretical framework of SEP, we could assume that mutualistic interactions are cases of reciprocal altruism. Accordingly, interspecific hunting associations among distantly related jackals and badgers, for example, would be considered equally costly to both parties in terms of energy or reproductive success. But assuming this is to leave unanswered the question of whether one or many of the organisms involved in a hunting association are actively modifying their social environment to benefit themselves more

than the other organisms, or to benefit the other organisms more than themselves. This presents a theoretical challenge because these associations can be interpreted in many ways. In any particular case, it will be difficult to determine whether an organism engaging in what looks like mutualistic interaction will be acting in favor of its own reproductive success, whether it is being deceived or manipulated by another organism to act to diminish its own reproductive success, or whether it is acting entirely altruistically to benefit another organisms' reproductive success at the moment or over the long term. Indeed, some cases will be cases of reciprocal altruism. But we cannot assume that they all are independent of empirical observation.

Furthermore, every case of mutualistic interspecific interaction creates a difficult *accounting problem* for SEP. How are energy expenditures during a long-term mutual interaction to be measured? Assuming this is possible, how are we to determine that the mental processes guiding the behaviours of these organisms are conforming to the rule of inclusive fitness or whether they are being influenced by environmental conditions that are the result of niche construction? For example, the niche constructing activities of organisms can have positive results on reproductive success. Often, however, these activities will cause the environment to become poorer in terms of energy resources. This in turn would put an energy strain on organisms in that environment and may result in a reduction of reproductive success. In both resource-rich and resource-poor environments, the factors that lead organisms to continue or cease modifying their environment through the construction of shelters, or migration, or by developing new associations, or breaking old alliances should be taken into consideration. NCT shows us that we cannot establish a priori that an organism's behaviour, especially its mutualistic interactions, conform to the rule of inclusive fitness. A detailed observation and analysis of an organisms' niche constructing activities and the effects of these activities will be required to more fully understand the selective processes that shape the psychological processes that are motivating its behaviour.

SEP's tenet of inclusive fitness focuses our attention on selective processes that influence the psychology of organisms from within. It is narrowly individualistic. By enlarging the theoretical framework to include NCT, we can examine the dynamic relationships

between organisms and their environment. This will allow us to identify otherwise ignored selective processes that shape the evolution of our psychology. I will have more to say about the implications of this theoretical enlargement on the explanatory method EEP below. For now, I will turn to the second proposed enlargement.

4.6.2 The mind is like a self-adaptive computer

According to the second tenet of SEP, the human brain's intelligence is like the intelligence embodied in a computer (Pinker 1997, p 26-27). SEP supports this tenet by appeal the *computational theory of mind*. As described in section 4.2.2 above, this theory holds that the mind operates according to classically computational principles such as the rule based manipulation of symbols. The similarity relation here being that the mind is like a computer in that it too manipulates its internal representations according to computational principles. The structure of this computational implementation is realized in the mind's complicated program. This program is coded in our genes as a result of evolutionary processes, and it manifests itself in the architecture of the mind which is a system of self-contained algorithms.

Human behaviour is subtle and flexible. And according to SEP, the subtlety and flexibility of human thought is the result of a complex algorithmic structure. This is why SEP posits the existence of hundreds or thousands of mental programs called modules. It is the complexity of modular programs and their interconnections—such as a program for vision and a program for snake avoidance—which results in the flexibility of behaviour. Each of these modules was selected for its causal contribution to solving a specific adaptive problem which occurred in during the Pleistocene.

Contrary to this view, I believe that the flexibility of behaviour and thought cannot be fully explained by appeal to the complexity of a fixed mental program. This is not to say that crucially important portions of human mental architecture are not fixed. I believe that certain mental programs have evolved as a results of natural selection and that we retain these portions partially due to genetic causal contribution. However, I believe that the flexibility of behaviour is made possible by evolutionarily fixed processes which enable, regulate, and maintain the flexibility of thought. This is a *self-adaptive computational*

theory of mind. In support of this theory, EEP like SEP, can maintain that the mind operates according to classical computational principles. However, unlike SEP, it must enlarge its theoretical framework to accommodate the claim that the mind is composed not just of static, albeit very complex, software, but of *self-adaptive software*. According to EEP, the mind is like a self-adaptive computer. It is composed of software that is designed by natural selection to be sensitive to its environmental goals and to its own internal processes. And it is designed to effect changes to those goals and internal processes. These claims draw support from the recent development of *self-adaptive software* [SAS] in computer science (Laddaga 2000; Ganek and Corbi 2003; Salehi and Tahbildari 2009). SAS is inspired by the realization that biological systems survive and reproduce in their changing environments by being sensitive to them and adapting or changing themselves in them.

At this point, we may pause and ask whether the mind's changing sensitivity and ability to adapt is not, as SEP argues, a product of a very complex but static internal program? If evolution occurred over a long period of time, surely the static program of the human mind reflects all of the evolutionarily significant contingencies that organisms need to survive and reproduce. The environment is complex, but so is the mind's software. Contrary to this however, SAS shows that SEP's exclusive focus on complexity is misplaced. The environment is indeed complex. But it is also constantly changing. As we saw in the previous section, organisms modify their environment, and they encounter ever changing environments as a result of migration, their interaction with other organisms, and as a result of the changes that other organisms have made to the environment. Complexity then is not the only important factor to be included in our understanding of the mind's architecture. Novelty by modification is key. If organisms have a mind that enables them to modify their environments, then it stands to reason that they also have a mind that enables them to modify their own mind in order to adapt to that environment. Self-adaptive software contributes to a theoretical framework for understanding how mind effects self-adaptive changes to itself in a changing environment.

The design of self-adaptive software started around the turn of the century. It was spurred by a Defense Advanced Research Projects Agency initiative seeking the creation of software that is capable of real-time self-regulation as opposed to preprogramed adaptation (Laddaga 2000). That is, rather than have programmers code *static* software into a computer, self-adaptive software projects have programmers code *self-variable* software processes. A computer running SAS is now able to modify both its internal *self-properties* and external *context-properties* (Salehi and Tahbildari 2009). In the former case, it does so via a real-time detailed description of its components (its modules). It is able to self-configure and reconfigure the arrangement and connections between modules (Taentzer et al. 2000). And it is able to effectuate fine grained changes to the adaptive algorithms internal to its modules (*Ibid.*). In the latter case of context-properties, SAS is aware of its environment (e.g. other computers on a network) and it is able to modify that environment by interacting in a cooperative or a self-protective manner (Ganek and Corbi 2003, p 6). This is facilitated by its ability to make predictions (*Ibid.*). SAS strives to implement these processes of internal and external sensation and change while hiding the complexity of these processes from the end computer user, and by aiming to decrease the number of changes effected from external programmers. It aims to develop a system that is both sensitive to its changing environment and autonomously modifiable. This accords with our understanding of the human mind. Although a detailed description of our mind's modular architecture may exist, we do not have access to it consciously. Thus, the mind hides the complexity of modular organizational and algorithmic modification from us (the end user). And the processes enabling such modifications do not have to be learnt. Rather, an important part of what is innately given to us by our genetic code are processes enabling us to adaptively self-modify our own minds.

Enlarging the theoretical framework of SEP allows us to notice that the mind's software is less like static software, and more like self-adaptive software. Our minds develop by means of *environmentally sensitive* processes and processes that enable us to *modify* ourselves and our environment. The most important and impressive adaptation that natural selection has designed is a mind that is able to achieve adaptations in real-time. This does not imply that the mind like a "blank slate"—a view staunchly admonished by

SEP. To say that the mind's software is self-adaptive software does not imply that there is no software "coded" in our genes. Nor does it imply, as Buller's criticism of SEP would have it, that the mind is entirely composed of *domain-general* as opposed to *domain-specific* (modular) processes. Modular processes exist. But their structure may not remain fixed throughout an organism's life. It is plausible that natural selection has also coded processes that regulate the modification of modular architecture and algorithms. As these modifying processes become better understood, they are becoming better implemented in computer science. Computer engineers posit the existence of *sensors* to monitor and reflect the state of a system's internal and external environment, and *effectors* which apply changes and realize adaptive actions in real-time (Salehi and Tahbildari 2009).

That the human mind can monitor its environment and significantly modify itself in order to adapt to that changing environment is important because evolutionary psychologists believe that the properties of particular modules depend on the specific function they carry out. On the other hand, SAS shows how the human mind likely modifies its modular properties in the face of novel functional problem requirements. This raises two difficulties for the view that the mind is composed of *static modules*. First, modular properties that are selected for the function that the module performed in the past are indeterminate. This is because a static modular account cannot tell the difference between those properties that may have been selected for their functional consequences in distant past (e.g. the Pleistocene) and those that may have been selected for their function in the more recent past (e.g. ancient Rome). The second difficulty is that modular properties that have remained fixed since the distant past because they served a specific function may have served different functions at different times. And perhaps more significantly, they may have served different functions throughout any particular organism's lifetime. The likely possibility that innate modular properties selected for one function are recruited for (or *exapted*) to perform a different function, or significantly modified to perform a completely novel function in light of changing environmental problem requirements is not explained on thestandard evolutionary psychological view that the mind's software is static rather than self-adaptive.

If we include SAS into the theoretical framework of enlarged evolutionary psychology, the extent to which human minds have modified themselves throughout evolutionary history and throughout any particular organism's lifetime becomes largely unknown. And though the evolutionary imperatives of survival and reproduction are surely limits on the selection for self-reflective and self-adaptive processes in past generations and past environments, a question about the speed of evolution arises. It is to this question that I turn next.

4.6.3 Evolution is faster than we thought: feminist evolutionists and biological leverage

As described in section 4.2.3 above, the third tenet of SEP is that evolution is slow. SEP holds that our evolved psychology—our basic preferences, motivations, and capacities—were selected for in the distant past or "deep time" (Buss 2010, p 3). And though the structure of our psychology is very complex, it is said to have remained fixed since deep time because evolution is "a glacially slow process that occurs in small increments over thousands and millions of years." (*Ibid.*) SEP holds that individual variations spread throughout the human population during the Pleistocene via inherited genes. This resulted in psychological traits that were universal to that population and are now universal to our population because they evolved during a very long period during which human ways of life are said to have remained the same (Barkow et al. 1992, p 5). The considered view of SEP is that there simply has not been enough time since the end of the Pleistocene for any significant evolutionary changes to our psychology to have occurred. However, contradictory evidence is mounting in evolutionary biology.

Many types of organisms exhibit evolutionarily significant morphological variations within short time scales. For example, the species-typical traits of Sockeye salmon have been documented as being directionally selected after just 13 generations (56 years) of isolated reproduction (Hendry et al. 2000). Likewise, notable body size increases have been documented in isolated populations of house sparrows in less than 100 generations (100 years) (Baker 1980). Empirical evidence of evolved physiological variation is also being published. Examples include salinity tolerance in copepods, heavy metal tolerance

in plants and animals, insecticide resistance in insects, and thermal tolerance in fish (Reznick and Ghalambor 2001, p 185). This accumulation of evidence is pointing towards the existence of a newly characterized phenomenon called "contemporary evolution" (Hendry and Kinnison 1999). Increasingly, studies are reporting that morphological and physiological evolution by natural selection occurs within short time scales of less than a few centuries (Hendry and Kinnison 1999; Reznick and Ghalambor 2001).

That evolution is more rapid than previously thought is gaining widespread empirical support across all of evolutionary theory. To better account for the rapid speed of evolution, EEP's theoretical framework enlarges to include *feminist evolutionary theory* and *biological leverage theory*. Evolutionary explanations in ecology, primatology, and evolutionary biology that make use of feminist evolutionary theory emphasize social organization and the influences of environmental variables (Haraway 1989; Hrdy 1997; Liesen 2007). This is a shift away from SEP's exclusive explanatory focus on processes internal to individuals. For example, SEP explains female sexual partner choice by appeal to psychological process leading them to "favor men who possess status and resources and to disfavor men who lack these assets." (Buss 1994, p 212) On the other hand, feminist evolutionary theory highlights *historical* and *situation-dependent* variables such as marriage systems and inheritance customs that enabled and supported the possession of wealth by males in medieval Europe (Hrdy 1997). At the very least, such historical and situation-dependent variables contribute to an explanation of why the psychological processes posited by SEP continue to be selected. They explain why selected psychological processes are sustained and become resistant to change. With regard to any particular behaviour under investigation, these variables by themselves may not explain why that behaviour occurs. This is not to say that genes play no causal role in causing psychological development and behaviour. But I think that including historical and situation-dependent variables is crucial to explain why a behaviour persisted in a particular time and place in the past (and perhaps why it persists today). It becomes an open question then whether historical and situational variables are causally responsible for the original occurrence and spread of the behaviour.

Historical and situation-dependent factors at least support the continued selection of innate mechanisms. Including them in explanation allows us to see that they influence the selection certain psychological processes as compared to others. Thus, feminist evolutionary theory reveals not only the possible structure of mechanisms internal to individuals, but the environmental structure that directs the natural selection of those mechanisms. Of particular note is experimental work demonstrating fast natural selection that is presently "ongoing in real-time" (Gowaty 2003; Gowaty et al. 2003; 2004; Moore and Moore 1999; Moore et al. 2003). Experiments on flies, cockroaches, and mice have been performed in which constraints on (female) organisms' choice of sexual partner have been manipulated. Those organisms who reproduced with partners of their choosing exhibited greater reproductive success than those that were constrained to mate with non-preferred partners. Their offspring were often greater in number and healthier. It is reasonable to infer then that organisms who were freer to mate with partners of their choosing were also likely to have had more and healthier offspring. What these experiments show is that current and past ecological constraints on the reproductive choices of organisms are evolutionarily significant. Thus, by including the perspective of feminist evolutionary theory, EEP is better equipped to "understand and test alternative hypotheses" for the origins of human psychology (Gowaty 2013, p 3).

Also with the aim of better accounting for rapid and on-going natural selection, EEP includes *biological leverage* in its enlarged theoretical framework (Barker 2008). Rapidly changing relationships among organisms can have significant evolutionary effects as in the case of mutualisms described in section 5.1. In such cases, an organism regulates the behaviour of another through open communication and indicators for coordinated action. In other cases, organisms can regulate another's behaviour in deceptive, manipulative, or coercive relationships. In such cases, an organism will regulate the behaviour of another for its own benefit. One way that organisms do this is by changing an *indicator* or by changing a *references state* which causally influences an adaptive process in another organism (*Ibid.*). For example, *lycaenid* butterfly larvae produce substances that attract ant workers and induce adaptive processes causing brood-care behaviour (Henning 1983). The butterfly larvae do this by providing the ants with a sweet honeydew-like

reward along with the chemicals that deceive their adaptive processes (*Ibid.*). This results in the butterfly larvae being carried back to the ants' nests (*Ibid.*). After some time in a nest, the butterfly larva acquire chemicals from the host ants which allows them to remain camouflaged while they prey on them. When the butterfly larvae enter the nest of the ants, they are redirecting (or *co-opting*) two of the regulative processes of the worker ants to their own regulative ends by presenting the ants with false indicators. The first of these is the process that regulates which organisms the ants care for. And the second regulates their communications. Once co-opted, these control systems can act as *biological levers* in that they are "a causal structure that transforms a small initial cause (acting on the reference or the indicator) into a much larger effect." (Barker 2008, p 12) The initial cause here (the butterfly larvae changing the indicator state of the ant larvae) is transformed into a much larger cause (widespread communication and helping behaviour). In turn, these large effects are highly consequential for the reproductive success of both the butterfly and the ants. They allow the butterflies to gain resources such as food and shelter. And they exert a selection pressure on the ants' communication processes, brood-care processes, and nest-building.

In human and primate cases, as in this insect example, dynamic and contested relationships of co-optive regulation by small initial changes can have widespread effects of great evolutionary importance. For example, as I discussed in my first paper, changes in whether two organisms share or perceive themselves to share the same values may change the *motivational orientation* they adopt towards each other. Doing so may also change whether an organism cares for another, and hence, whether an organism empathizes with that other organism. The parallels between the biological levers in the above ant and larvae case and possible human cases of mimicry, deception, and exploitation in contested zones are striking. I will elaborate on some of these parallels towards the end of this paper when I discuss the role that biological levers play in the new methodology of EEP and when I sketch an alternative EEP explanation of how empathy evolved. For now, I will turn to showing how the tenets of EEP's enlarged framework are mutually supportive, and then to drawing out the implications that this framework has for the revised method of EEP.

4.7 A mutually supportive theoretical enlargement

The three tenets of standard evolutionary psychology are:

1) Inclusive fitness

2) Computational theory of mind

3) Slow evolution

As the name suggests, enlarged evolutionary psychology enlarges each of these tenets to include:

4) Niche construction and mutualism

5) Self-adaptive software

6) Feminist evolutionary theory and biological leverage

I take the EEP to be providing a theoretical enlargement of SEP. But with regard to some of EEP's theoretical tenets, "enlargement" does not only imply extending SEP's meta-theory. EEP's enlargement also requires rejecting some of SEP's deeply held befliefs. No element of the enlarged theoretical framework of SEP necessarily implies the falsity of the original three tenets of SEP. But all elements of the enlarged theoretical framework imply that the three tenets of SEP's original framework cannot be maintained exclusively. For example, SEP can no longer hold that evolution is always very slow given the recent empirical evidence from feminist evolutionists and the theoretical support from biological leverage. The tenet that evolution is very slow cannot be maintained to the exclusion that evolution can occur over much smaller time scales. To take another example, the inclusion of niche construction theory does not imply that we do not sometimes (or often) behave in ways that favour the reproductive success of targets with whom we share the most genes. However, it does allow us to form different hypotheses about why some of our behaviour often conforms to inclusive fitness theory. It may turn out that in many cases we have constructed our environment in ways that directs our behaviour according to the predictions of inclusive fitness theory. That the

majority of parents in North America live together with their children in shelters that territorially exclude other parents and other children is not merely because our psychological mechanisms aim towards the maximization of inclusive fitness. Rather, it is because, in North America, we have constructed and continue to construct our shelters and our family relationships in such ways. Different shelters and family arrangements have been the norm in North American history and continue to exist elsewhere in the world. Humans, like other organisms, have constantly modified their environments. And, in turn, we have constantly adapted to our modified environments. If we had not, then we would likely not have survived to the present today.

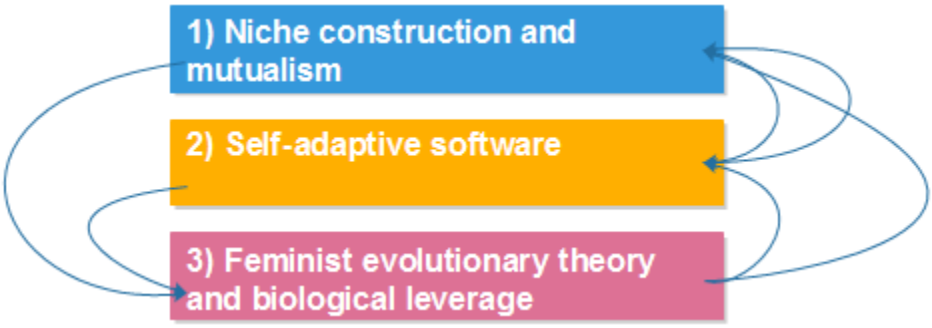
We construct our environment, and our environment influences our behaviour. But our environment also influences the evolution of our psychology. More specifically, the creativity and variety of our environmental modifications and the fact that we have adapted to environmental modifications are reasons to believe the mind is more like a self-adaptive computer than a computer whose program has been fixed for a very long time. The inclusion of *niche construction theory* into the theoretical framework of evolutionary psychology both supports and is supported by a *self-adaptive computational theory of mind*. This theory states that the flexibility of human behaviour is made possible by the flexibility of human mental structures. Throughout life, the human mind, like a self-adaptive computer, effects changes to its goals and internal psychological algorithms or processes as a function of the adaptive problems it faces in its environment. This again is not to deny that certain fixed structures of the mind continue to be selected because they contributed to the solution of recurring adaptive problems in the past (as would follow from SEP's tenet that the mind is like a classical computer with fixed algorithms). Natural selection continues to select for certain fixed structures that play a role in the processes of visual perception. But this does not imply that vision is and always has been the same. EEP denies the claim that all of our psychological processes have been developmentally fixed since the Pleistocene. Clearly, the mind's processes of visual perception are self-modified throughout one human lifetime. An infant does not see in the same way as an adult. Furthermore, a painter may not see in the same way as before they began to paint. I think this provides some reason to believe that even the processes of

visual perception have evolved throughout human history. Humans today may perceive some elements of the environment in the same way as humans who lived during the Pleistocene. But if the mind is like a self-adaptive computer and not a pre-adapted computer, it seems increasingly likely that we see differently from humans who lived in the distant past.

But for now, this is an aside. The important point is that, when taken together, niche construction theory and self-adaptive software provide us with a new ways of describing how human minds evolve. Specifically, they evolve in virtue of the natural selection of processes enabling sensitivity to and modification of environments (including the mind's own psychological architecture or environment). Such a description urges us to accept that evolution by natural selection operates much faster than we previously thought. And if we accept that evolution is rapid and ongoing, then this supports the inclusion of both feminist evolutionary theory and biological leverage theory. The former integrates social and historical factors into our evolutionary explanations, whereas the latter enables us to describe regulative interactions that are otherwise difficult to formalize, such as mutualism, mimicry, deception, and manipulation.

The three enlarged tenets of EEP are mutually supportive in a way that makes holding SEP's three tenets to the exclusion EEP's enlarged tenets inconsistent. And EEP like SEP, provides a unified theoretical framework for explaining human behaviour as a function of biological processes; some of which were selected in the past, and some of which are subject to ongoing evolution by natural selection. But this new and enlarged framework also adds a lot of complexity to evolutionary psychological explanations. While, as mentioned, this added complexity does not necessarily negate any of the previously held tenets of SEP, it does present important challenges to its methodological assumptions.

Figure 5: *Mutually supportive tenets of EEP*



In the next section I elaborate these challenges, and present a revision of SEP's explanatory method that accords with the enlarged theoretical framework of EEP. I then conclude the paper by generating a hypothesis about the evolution of empathy from this revised method.

4.8 The revised method of EEP

1) As outlined in section 4.4, the explanatory method of standard evolutionary psychology —the way it generates and tests hypotheses about contemporary phenomena—is as follows:Choose a widespread phenomenon.

2) Posit innate, universal, and historically fixed mechanisms that cause the phenomenon in (1).

3) Generate and test hypotheses about how the mechanisms in (2) functioned to cause behaviour that solved recurring adaptive problems during the Pleistocene.

4) Justify the existence of the mechanisms in (2) by appeal to statistical normality in (1) and historical plausibility in (3).

The method of EEP leaves the first step as is. This is because SEP and EEP share the same main explanatory goal: to explain widespread phenomena in terms of the causal processes accounting for what exists (Buss 1996, p 297). However, the enlarged

theoretical framework of EEP does suggest a revision of the second step of SEP's method.

SEP looks to explain the cause of the phenomenon under investigation in (1) in terms of a specific type of causal process, namely, psychological processes. This is well and good. Psychology is and should be about accounting for what psychological processes exist, and accounting for how they work. And, as discussed above, the psychological processes posited by SEP are those that are innate because they solved adaptive problems humans faced in the distant past. The reason why SEP posits *innate* psychological processes is because it holds that they were selected in the past because of the consequences they functioned to produce. These evolutionary functions reveal the form of the mind today:

> In evolved systems, form follows function. The physical structure is there because it embodies a set of programs; the programs are there because they solved a particular problem in the past. (Cosmides and Tooby 1997, p. 13)

"Form follows function" where 'function' refers to the consequences a process functioned to produce in the past, and 'form' refers to the stimuli the process is triggered by and its algorithmic structure in the present. Accordingly, in step (2), SEP posits the existence of processes that "follow" the "function" of solving adaptive problems.

Like SEP, EEP allows for form to follow function. But, importantly, it holds that form also *assists* function. The forms or processes of our minds are there because they performed a function in the past. But our psychological processes also motivated us to modify the environments in which we lived in order to contribute to the functions that our minds were selected to perform. We found and made shelters, tools, and clothing. We created abstract social relationships such as clans, tribes, and marriage. We regulated each other's goals and preferences in order to achieve co-operative or competitive benefits. Such environmental modifications, along with psychological processes, causally contributed the functions that our minds performed in order for us to survive and reproduce.[52] In turn, these environmental modifications affect the algorithms of our mind

[52] Sterelny calls such reproductive successes-assisting environmental modifications "scaffolding" (Sterelny 2003). In the extended mind debate, they are called "spreading the cognitive load" or

in virtue of the mind's self-adaptive nature. That is, environmental modifications have been produced by the psychological processes that enable them. Thus, the functions that we needed to perform in order to survive and reproduce have changed over evolutionary time. So too have the psychological processes that causally contribute to those functions. The goal here for EEP will be to clearly identify when there has been (or could be) a change in the process or underlying mechanisms as opposed to merely a change in realization of the same process type. To this end, existing research on the self-modifying capabilities of self-adaptive software may elucidate the self-modifying capacities of the human mind. In the self-adaptive software literature, self-modifications processes include *distributed graph transformations* (Taentzer et al. 2000), many types of *reflective middleware* (Huang 2003), and *architecture description* languages (Salehi and Tahbildari 2009, p 13).

Provisionally, however, if we accept that mind modifies its environment and itself to assist in performing evolutionary functions, then we should first seek to explain any widespread phenomenon under investigation in terms of the psychological processes that *follow from and assist* the functions that these processes perform in the *contemporary environment*. This is a crucially important step if we are to understand why and how (if in any way) the functions of our psychological processes have evolved.

This challenges the assumption in step (2) that the mind is composed of functional processes that have remained fixed. Accordingly, the revised method of EEP does not begin by positing the existence of processes that have remained fixed because of their functions during Pleistocene. Again, this is not to say that such historically fixed processes do not exist. But within the theoretical framework of EEP, rather than assuming their existence, they must be discovered. To discover whether a process has remained fixed since ancient time then, we should begin by distinguishing the possible ancient functions of such a process from its possible contemporary functions. This will involve identifying the *causal feedback loops* or dynamic causal links between the causes and effects of a biological phenomenon (Oyama 2000; Oyama et al. 2001). In other

---

"cognitive off-loading" (Clark and Chalmers 1998; Menary 2007; Clark 2008; Wilson and Clark 2009).

words, it will involve identifying the functions of the phenomenon by the many roles it plays in a specific environment thereby contributing to the maintenance of that environment's continuity or change. Insofar as a psychological process causally contributes to a phenomenon under investigation, the question to ask then becomes: what is the function of the psychological processes causing the phenomenon in terms of its consequences that either sustain or change the selection of organisms in the environment? Thus, the revised method of EEP replaces step (2) of the standard method with:

2) Describe the current functions of the phenomenon in (1).

As just mentioned, an explanatorily fruitful way of describing the contemporary function of a psychological process is in terms of the causal feedback loops it participates in. These feedback loops either contribute to sustaining or changing the environment in which organisms are selected. This way of characterizing the present functions of process allows for it to appear in a complex environmental context and as part of the larger dynamic systems of which it is a part (Jax 2005, p 1). This way of describing a process's *current functions* is almost the same way that SEP describes the *past functions* of the psychological processes it posits. The notions of *function* and the means of identifying a processes' functional consequences are available to both methods. The only significant difference is the environment being referred to. Whereas in step (2) of SEP's explanatory method, the environment being referred to is that of the Pleistocene, in step (2) of EEP's revised method, the environment being referred to is the contemporary environment. This bring us to step (3).

Step (3) of SEP is to generate hypotheses about the functions of psychological processes causing a particular phenomenon in the environment of the Pleistocene. The method of EEP leaves this step as is. The question of this third step then is: what are the functions of the psychological processes that contributed to behaviours which solved recurring adaptive problems during the period approximately 1.8 million years ago to 13 000 years ago? Identifying the adaptive problems that humans faced during this period requires identifying the "adaptation-relevant properties of the ancestral environments encountered

by members of ancestral populations" (Tooby and Cosmides 1990, p 386). These adaptation-relevant properties or statistical "invariances" are properties that can be "described as sets of conditionals of any degree of complexity, from the very simple… to any degree of conditional and structural complexity that is reflected in the adaptation." (Tooby and Cosmides 1990, p 389). And the way to identify these properties is by inferring them from the present (Tooby and Cosmides 1990, p 390). For example, "[t]he presence of psychological mechanisms producing male sexual jealousy tells one that female infidelity was part of the human and ring dove EEAs [environment of evolutionary adaptedness]." (*Ibid.*)

Many critics of SEP have presented a variety arguments for why describing the environment of the Pleistocene is epistemologically difficult (Stotz and Griffiths 2002; Sterelny 2003; Buller 2005; Griffiths 2007; Liesen 2007). Some of these arguments are more compelling than others. And as mentioned, the enlarged framework of EEP which includes niche-construction theory, feminist evolutionary theory, and biological leverage presents challenges to the tenet that all innate human psychological processes motivate us to maximize reproductive success. But I will not rehearse these arguments at present because the revised method of EEP, like that of SEP, makes room for the possibility that many of our psychological processes are innate and that they contributed to the performance of evolutionarily significant functions in the distant past.[53] Thus, the third step is:

> 3) *Generate hypotheses about the functions of the behaviours in (1) during the Pleistocene and posit that the psychological processes causing them are innate.*

This third step follows the second step in an explanatorily beneficial way. This is because SEP specifies that hypotheses about the functions of behaviours during the Pleistocene should be generated from information available in the present, which is precisely what

[53] This is not to say that these processes are innate *because* they performed an evolutionarily significant function in the *distant* past. If we accept that evolution is fast, then it is not unlikely that innate psychological processes continue to be selected because of their functions in the *nearer* past.

EEP's step (2) provides. The information gathered in step (2) (specifying the current functions of a behaviour) serves as a base for the generation of hypotheses about their functions in the distant past. These hypotheses, as in the case of male sexual jealousy above, may take the following form: If psychological processes contribute to causing X in the current environment, and the function of X in the current environment is Y, then X may have been present in the EEA. Thus, the question of this step remains: what was the function of X in the EEA? However, the revised method of EEP allows us to immediately notice whether the function of X (the phenomenon under investigation) was the same in the current environment as it was in the EEA. This is an improvement on SEP's method because whether the function of a psychological process causing a phenomenon has remained fixed since the Pleistocene can now be discovered and evaluated rather than assumed. Evaluating this possibility was previously impossible without a functional description of the phenomenon in the current environment being part of the explanatory method. We can now in step (4) compare the function posited in (2) with that posited in (3) to identify whether they are the same or not:

4) *Compare the functions in (2) with those in (3).*

This brings us to the step (5).

In the two last steps we generated hypotheses about the functions of a phenomenon caused by psychological processes in both the current environment of observation and the distant past environment. And we compared those functions to identify whether there were any in common. This allows us to examine whether the functions of the psychological processes causing a phenomenon will be different in the present as compared to in the past. For example, the function of male sexual jealousy in the current environment may turn out to be different from the function of male sexual jealousy in the distant past environment of the Pleistocene. This presents a challenge in the final step of SEP's explanatory method. In the final step of SEP's method, the existence of an innate psychological processes was justified by appeal to both the statistical regularity of the phenomenon it caused in the current environment and the historical plausibility of it

causing the same phenomenon in the distant past. But as we have just seen, the same phenomenon (caused by psychological processes) will sometimes have different functions in these two environments. Why is this the case? SEP's answer is that the environment has become much more complex, and that it is this environmental complexity that has changed the function of the phenomenon (and hence the psychological processes causing it). According to SEP, even though the environment changed, our psychology stayed the same. On the other hand, the enlarged theoretical framework of EEP allows us to answer differently. EEP proposes:

> 5) *Track similarities and differences between current functions and historical functions and identify possible correlations with significant environmental changes or biological levers.*

The revised method of EEP does not assume that a current behaviour can be explained only in terms of the psychological processes that cause it. Such processes are integral causal components, but EEP views the continued existence of a behaviour as the outcome of psychological processes and its current functional consequences. The same goes for explaining a phenomenon in the distant past. The phenomenon must be explained by positing psychological processes that caused behaviours whose consequences enabled their continued selection. If form (psychological processes) follows function (evolutionarily significant consequences), then the question becomes: what were the functions that our psychological processes contributed to throughout evolutionary history? This is the fundamental question of any evolutionary psychology.

To answer this question, step (5) of EEP's method proposes that we track how a phenomenon's function has changed over time (if it has changed), and correlate those functional changes with significant environmental modifications or biological levers.[54] It is likely that in many cases, it will be impossible to carry out this historical tracking in detail. But in successful cases, the method specifies that if a functional change can be

[54] Environmental changes will be evolutionarily significant when they are caused by niche-construction, social arrangements, or other natural changes, such as disasters or changes in climate that sharply affect reproductive success.

correlated with a significant environmental change or biological lever, then these correlations are evidence for possible changes to a population's psychological processes over time. A change in a phenomenon's function that is sustained over several generations by a change in the environment or that is having large scale effects due to biological leverage may be evidence of a significant change in our psychology. Accordingly, the comparative method of EEP allows for evolutionary changes in human psychology to be discovered. Rather than assuming that behaviour is caused by an unchanging structure, EEP allows us to track when and how a structure that was fixed in the past may have changed and whether it has become fixed again.

This brings us to the last step of EEP's comparative method:

6) *Justify the existence of the distant historical psychological processes in (3) by appeal to statistical normality in (2) and significant evolutionary changes or continuities in (5).*

In this final step, we can assess whether the distant historical functions (and form) of the psychological processes we posited at the outset have continually been selected (i.e. have remained fixed) or whether they have changed. A strength of this comparative method is that it allows us to provide evidence for when we expect a functional change to be accompanied by a change in psychological form (or algorithmic structure) as opposed to merely being the result of increased environmental complexity. As mentioned above, this is because in step (5) we will have correlated possible functional changes with environmental changes and inter-organismic psychological regulations that will likely have had important effects on reproductive success (e.g. niche-construction, biological levers).

To summarize, the explanatory methods of SEP and EEP are as follows:

Table 2. *The method of SEP and EEP*

Standard Evolutionary Psychology:	**Enlarged Evolutionary Psychology:**
1) Choose a widespread behavioural phenomenon.	1) Choose a widespread behavioural phenomenon.
2) Posit innate, universal, and historically fixed processes whose past function causes the phenomenon in (1).	2) Describe the current functions of the phenomenon in (1).
3) Generate and test hypotheses about how the mechanisms in (2) functioned to cause behaviour that solved recurring adaptive problems during the Pleistocene.	3) Generate hypotheses about the functions of the behaviours in (1) during the Pleistocene and posit that the processes causing them are innate.
4) Justify the existence of the processes in (2) by appeal to statistical normality in (1) and historical plausibility in (3).	4) Compare the functions in (2) with those in (3).
	5) Track similarities and differences between current functions and historical functions and identify possible correlations with significant environmental changes or biological levers.
	6) Justify the existence of the distant historical processes in (3) by appeal to statistical normality in (2) and significant evolutionary changes or continuities in (5).

The two most significant revisions to the method of SEP proposed by the comparative method of EEP are step (2) (the description of current functions) and step (6) (ontological justification by appeal to the evolutionarily informed historical comparisons). Step (2) is one which has not been entirely ignored by SEP, but it is one which has heretofore not been integrated into its explanatory method. Often, evolutionary psychologists will describe the function of a phenomenon caused by current psychological processes. But their theoretical assumptions then very quickly lead them to posit innate processes that have remained fixed since the unspecified distant past to explain this current function, despite the fact that it functioned differently in the past.

The significance of step (6) is that it allows us to preserve some of the universal ontology of SEP while also allowing us to discover if and when that ontology became universal and how it might have spread. If some of the current functions of a phenomenon under investigation have remained the same since the very distant past (e.g. the Pleistocene), then this will provide evidence which can be used to justify the claim that this phenomenon is caused by an innate psychological process. However, if the current functions of a phenomenon are very different from those that we hypothesize were its functions in the distant past, then this will be a clue for us to investigate and follow the comparative steps of EEP. We can examine this clue (the difference between current functions and distant historical functions) and search for more recent historical factors that may have led to widespread selective variation across a population. Given the more recent occurrence of these factors, they should be correlated with environmental changes or biological levers affecting selection. This will then provide evidence which can be used to justify the claim that a phenomenon is caused by an innate psychological processes in a given population. At that point, developmental and cross-cultural analysis could be performed to assess whether the phenomenon—e.g. marriage—is universally caused by innate psychological processes.

The comparative method of EEP allows for new possibilities of discovering how our minds have evolved. But it also raises many questions that have not been sufficiently addressed by evolutionary psychology. Such questions include: how does a change in psychological process becomes reliably heritable? Under what conditions and how long does it take for a psychological

change via biological leverage to spread across a population? If, over many generations, an environmental stimulus results in a psychological change across processes in more than one specific domain, what implications does this have for explanations that appeal to a single domain specific processes causing our responses to such stimuli? Answers to these questions about the explanatory method of EEP and other questions about its theoretical framework will require further research.

4.9 Enlarged evolutionary psychology of empathy

As we saw in section 4.1, SEP posits innate psychological processes that cause females to empathize more frequently and more accurately than males. It claims females possess these processes because the consequences of empathy—its functions—were mostly caused by females in the environment of the Pleistocene. As mentioned earlier, these functions include: 1) "mothering"; 2) "gossip"; 3) "social mobility"; 4) "reading your partner". The empirical findings of SEP support the claims that, in today's environment, females are on universal average more sensitive than males to the needs of their children; that females are more likely to gossip amongst themselves than males; that females usually have to adapt to living with the extended family of their male partners; and that females are better at becoming aware of the needs of their male partners. According to SEP, these are some of the contemporary functions of empathy, and they explain essential differences in the occurrence of empathy because these functions contributed to the reproductive success of a population over a long period in the distant past.

In accordance with step (2) of EEP's explanatory method, we can accept these four current functions of empathy. However, in step (3), it is possible to infer from these current functions that empathy would have been equally important to the reproductive success of males as it was to that of females. As I have argued in my first paper, empathy is more likely to occur in a *cooperative motivational orientation*. Hunting is a paradigmatic example of such a behaviour which is often assumed to have been performed mostly by males. It is a reasonable then to generate the following hypothesis: that one of the functions of empathy in the Pleistocene was to favour cooperation (reading one's partner) and helping behaviour while hunting. Males may have benefited from becoming aware of the concerns of their hunting associates. These empathic hunters would have been keenly motivated to pay attention to and help their injured associates

and been able to solicit help in dangerous circumstances once injured. When we compare the current functions of empathy with its possible historical functions (step 4), we notice that not all of the plausible historical functions of empathy are supported by current empirical findings in SEP. To see why this is the case, we must proceed to step (5).

In this step we attempt to track the similarities and difference between the current functions and historical functions of empathy. And then we attempt to correlate these with significant environmental changes or biological levers. So far, we can see that the items in the set of current functions of empathy that we have posited are very similar to those in the set of historical functions. *Mothering*, *gossip*, *social mobility*, and *reading your partner* are both current functions and distant historical functions. This does not mean that the set of current functions is full. There are likely to be many other functions of empathy in the contemporary environment. For example, one of the current functions of empathy is motivating people to give money to charities. And there are likely to be many additional distant historical functions that we have not yet discovered. But let us stay with our current addition to the previous four distant historical functions, namely hunting.

Can a change in empathy's hunting function be correlated with an evolutionarily significant environmental change or biological lever? One such change is the development of agriculture, approximately 520 generations ago. Agriculture can be considered a form of niche-construction that possibly evolved from *forest gardening*—the practice of identifying and protecting certain desirable and useful plants. As populations increasingly modified their environment in order to regulate consumable plant and animal growth, males shifted away from hunting. Whereas hunting required people to quickly become aware of their associates' concerns and to respond rapidly with helping behaviour when appropriate, agriculture was a much slower and safer activity. However, a psychological effect of increased agriculture may have been decreased male empathy. As hunting decreased, so too did the hunting function of empathy. This correlation between a change in empathy's function and a significant environmental change due to niche-construction could serve as evidence that agriculture resulted in population-wide psychological changes that we have hitherto not discovered. For example, it may be that empathy is more frequent and accurate among populations that still practice hunting and gathering regularly and

that this is partially caused by a difference in psychological processes. This is a question that cannot be answered at the moment for lack of empirical evidence. But the weaker claim that agriculture led to a decrease in the degree of male empathy can be defended on historical grounds.

This claim can be supported by correlating the decrease of empathy's role in male hunting with a *biological lever*. Agriculture had many effects on organism relationships, including a change in the relationships between males and females. Prinz notes that early farming equipment was hefty and operated mostly by males (Prinz 2007, p 298). He argues that this led to male control of wealth, commerce, and later literacy, and political power (*Ibid*). This account of how agriculture affected the dynamic interactions of males and females is plausible, but I think lacking in some key details. Even if we provisionally accept Prinz' account, it seems unlikely that the control of resources, education, and political power by one subgrouping of the population was acquired while the rest of the population acquiesced. After agriculture, resources and social arrangements took on new dimensions of contestation. To gain and maintain control of these resources in this new environment, males may have acted so as to cause females to be more sensitive to the concerns of males. One way that this may have been achieved is by a change in the *regulated indicator* of who to empathize most often with, namely to adult males. In biological terms, let us assume that females possessed evolved psychological processes whose functional goal was regulate empathy with other organisms (Barker 2008, p 10). The *indicator value* which was used to sense whether and with whom empathy was occurring may have been, as SEP suggests, one which favoured certain organisms rather others. For instance, females may have up to that point evolved to regulate empathy by checking whether they were empathizing with young humans more often (and perhaps more accurately) than with older humans. The indicator in this evolved psychological process would have been open to both co-operative and exploitive co-optation (Barker 2008, p22). In this evolutionary scenario, certain males may have succeeded in modifying the behavioural indicators of females through teaching, modeling, aggression, or other means. Accordingly, modifying the psychological process of empathy may have been a biological lever. This small change in the evolved structure female empathic psychological processing would have led to much larger effects. Once the indicator change was *effected*, not empathizing with males more often or more accurately would have been taken by females to be

an error. In this way, the concerns of males and female motivation to help males with their concerns would become more evolutionarily significant. This small indicator change alone would have affected the dynamic behavioural interactions of males and females, their social roles, the types of institutions and environmental structures they created, the distribution of resources, system of laws etc. The scope of these possible causal feedback loops are beyond the scope of the present analysis. It remains an important and complex question whether in populations which regularly hunt, gather, and forest garden, females are more, less, or equally empathic as compared to males. But if a biological lever leading to increased female empathy for the concerns of males can be correlated with the development of agriculture (an environmental change), then according to EEP, this is a place to look for psychological modifications (and self-modifications) that may have had large-scale effects and that may now be part of our innate human psychology.

This brings us to the final step of EEP's method: justify the existence of psychological processes in females causing more frequent and accurate empathy in a specific population by appeal to contemporary empirical findings in support of the functions of empathy in the current environment, and two significant evolutionary changes—namely, the niche-constructing activity of agriculture, and the biological leverage of the indicator value of empathy determining potential targets of empathy. The former may have had the effect of decreased empathy in males, and the latter increased empathy by female agents for male targets. We can also predict that these differences between male and female empathy will not be as great as they could be because, as we have seen, many of the current functions of empathy have remained the same since the Pleistocene. However, it is important to note that these functions have also undergone significant evolutionary changes during certain times.

The implication of this for Baron-Cohen and Pinker's accounts of empathy is that their relatively restricted meta-theory results in a method that is potentially inductively weaker than it purports to be. There are variables that can more precisely explain the evolution of empathy that have not been justifiably eliminated. As I have sketched above, these variables (e.g. niche construction, psychological self-adaptation, biological levers) can be applied. But to know whether doing so will yield empirical results that will support my sketch, we must await experimentation. In the

meantime, however, I believe that future experimentation on the evolutionary psychology of empathy should adopt the enlarged perspective that I have argued for if only to determine whether SEP, as it has been practiced up now, is as inductively strong as it purports to be. Empathy's psychological functions of childcare, education, communication, social organization, and awareness of the concerns of others could all possibly be better explained by the comparative method of EEP. I hope to have shown that by enlarging the theoretical framework and revising the method of standard evolutionary psychology, we will be able to more precisely identify the evolved psychological traits of contemporary human beings while more accurately explaining why, when, and how they evolved.

4.10 Concluding remarks

The most influential accounts of how and why empathy evolved come from *standard evolutionary psychology.* These accounts have yielded impressive results and have gained widespread acceptance because they are supported by a coherent evolutionary meta-theory and an established methodology. This methodology leads SEP researchers seeking to explain contemporary behaviour to posit the existence of psychological processes that have remained functionally fixed since Pleistocene. But doing so excludes the possibility of more recent evolutionary changes for which there is increasing theoretical support. Moreover, SEP presents a view of the mind according to which it is adapted to a specific environment in the distant past. The implications for the evolution of empathy are clear. Its current functions have remained fixed, and its implementation is aimed at conforming to historical imperatives of an environment in which there were *recurrent adaptive problems* and *fixed social roles*.

In this paper I have argued for an alternative explanation of empathy's evolution. This explanation is informed by an enlarged evolutionary meta-theory and a revised methodology, which together I have called *enlarged evolutionary psychology.* EEP integrates the recent theoretical developments of *niche construction theory and mutualism, self-adaptive software, feminist evolutionary theory,* and *biological leverage.* This enlargement allowed me develop a new *comparative method* for EEP. According to this comparative method, a researcher seeking to explain contemporary behaviour must first develop a hypothesis about the contemporary functions of the psychological processes that the behaviour is partially caused by. The researcher

can then compare these possible contemporary functions to the possible distant historical functions and attempt to track changes in the operation of an evolved psychological process by correlating relevant environmental changes and modifications with biological levers and other possible means of psychological self-adaptation. Then I sketched how EEP can be used to provide an alternative explanation of empathy's evolution according to which there may have been significant evolutionary changes in our psychological processes that have occurred since the development of agriculture. This alternative evolutionary explanation of empathy is consistent with the care account of empathy presented in my first two papers. The explanatory role of changes in environmental contexts is supported by the inclusion of niche construction theory. And the self-modification and co-regulation of values and motivations is supported by the inclusion of self-adaptive software, feminist evolutionary theory, and biological levers. In this way, the care account better fits the theoretical framework of EEP. However, the alternative explanation here sketched supports but does not establish the truth of the care account. Nor does it establish the truth of EEP in opposition to SEP. However, the framework and method of EEP is at odds with that of SEP because it posits that human minds modify themselves by modifying their environment and by using their self-adaptive capacities, which in turn can result in large-scale functional changes to those processes over short time periods. It is a significant consequence of this that SEP may, to the peril of its objectives, be neglecting significant evolutionary changes in our psychological processes and much of the complexity of how and why empathy evolved.

References

Adams, Frederick. “The Informational Turn in Philosophy.” *Minds and Machines* 13.4 (2003): 471–501. Print.

Baker, Allan J. “Morphometric Differentiation in New Zealand Populations of the House Sparrow (Passer Domesticus).” *Evolution* (1980): 638–653. Print.

Barker, Gillian. “Biological Levers and Extended Adaptationism.” *Biology & Philosophy* 23.1 (2008): 1–25. Print.

Barker, Gillian, Eric Desjardins, and Trevor Pearce, eds. *Entangled Life, Organism and Environment in the Biological and Social Sciences*. Dordrecht, Heidelberg, New York, London: Springer, 2014. Print.

Barkow, Jerome H., Leda Ed Cosmides, and John Ed Tooby. *The Adapted Mind: Evolutionary Psychology and the Generation of Culture.* Oxford University Press, 1992.

Baron-Cohen, Simon. *Essential Difference: Male and Female Brains and the Truth about Autism*. Basic Books, 2003.

---. “The Evolution of Empathizing and Systemizing: Assortative Mating of Two Strong Systemizers and the Cause of Autism.” *The Oxford Handbook of Evolutionary Psychology*. Ed. R.I.M. Dunbar and Louise Barrett. Oxford University Press, 2007. 213–226. Print.

---. “The Male Condition.” *New York Times* 8 Aug. 2005: Print.

---. “The Science of Evil.” *New York Times* a 2011: Print.

---. *The Science of Evil: On Empathy and the Origins of Cruelty*. Basic books, 2011.

Baron-Cohen, Simon, Rebecca C. Knickmeyer, and Matthew K. Belmonte. “Sex Differences in the Brain: Implications for Explaining Autism.” *Science* 310.5749 (2005): 819–823. Print.

Baron-Cohen, Simon, and Sally Wheelwright. “The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences.” *Journal of autism and developmental disorders* 34.2 (2004): 163–175. Print.

Batson, C. Daniel. *The Altruism Question: Toward a Social-Psychological Answer*. Psychology Press, 1991.

Berkeley, István SN. “Some Myths of Connectionism.” (1997).

Brewer, Marilynn, and Linda Caporael. “An Evolutionary Perspective on Social Identity: Revisiting Groups.” *Evolution and Social Psychology*. Ed. Mark Schaller, Jeffry Simpson, and Douglas Kenrick. Psychology Press, 2013. Print.

Brown, Rachael. “Rethinking Behavioral Evolution.” *Entangled Life*. Springer, 2014. 237–260.

Bshary, Redouan et al. “Interspecific Communicative and Coordinated Hunting between Groupers and Giant Moray Eels in the Red Sea.” *PLoS biology* 4.12 (2006): e431. Print.

Buller, David J. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. MIT Press, 2005.

Buss, David. “Sexual Conflict: Evolutionary Insights into Feminism and the ‘battle of the Sexes.’” *Sex, Power, Conflict: Evolutionary and Feminist Perspectives*. Ed. David Buss and Neil M. Malamuth. Oxford University Press, 1996. 296–318.

---. “Why Students Love Evolutionary Psychology... And How to Teach It.” *American Psychological Association Education Directorate* 20.10 (2010): Print.

Buss, David M. “Evolutionary Psychology: A New Paradigm for Psychological Science.” *Psychological inquiry* 6.1 (1995): 1–30. Print.

---. *Evolutionary Psychology: The New Science of the Mind.* Allyn & Bacon, 2008.

---. *The Evolution of Desire: Strategies of Human Mating.* Basic books, 1994.

---. *The Handbook of Evolutionary Psychology*. John Wiley & Sons, 2005.

Buss, David Michael, and David P. Schmitt. “Evolutionary Psychology and Feminism.” *Sex Roles* 64.9-10 (2011): 768–787. Print.

Buss, David M., and Neil M. Malamuth. *Sex, Power, Conflict: Evolutionary and Feminist Perspectives.* Oxford University Press, 1996.

Chalmers, David. “Why Fodor and Pylyshyn Were Wrong: The Simplest Refutation.” *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society, Cambridge, Mass*. 1990. 340–347.

Chalmers, David J. “Connectionism and Compositionality: Why Fodor and Pylyshyn Were Wrong.” (1993).

Clark, Andy. *Associative Engines: Connectionism, Concepts, and Representational Change*. MIT Press, 1993.

Clark, Andy, and David Chalmers. “The Extended Mind.” *analysis* (1998): 7–19. Print.

Clutton-Brock, Tim. “Cooperation between Non-Kin in Animal Societies.” *Nature* 462.7269 (2009): 51–57. Print.

Cosmides, Leda, and John Tooby. “Evolutionary Psychology and the Emotions.” *Handbook of emotions* (2000): 91–115. Print.

---. “Evolutionary Psychology: A Primer.” *Evolutionary Psychology: a primer* (1997).

Dawkins, Richard. *The Selfish Gene. New York*. Vol. 1. 1976. Print.

De Waal, Frans BM. *Good Natured*. Harvard University Press, 1996.

Dickman, Chris R. “Commensal and Mutualistic Interactions among Terrestrial Vertebrates.” *Trends in ecology & evolution* 7.6 (1992): 194–197. Print.

Fodor, Jerry A. “Why There Still Has to Be a Language of Thought.” *Computers, Brains and Minds*. Springer, 1989. 23–46.

Fodor, Jerry A., and Zenon W. Pylyshyn. “Connectionism and Cognitive Architecture: A Critical Analysis.” *Cognition* 28.1 (1988): 3–71. Print.

Ganek, Alan G., and Thomas A. Corbi. “The Dawning of the Autonomic Computing Era.” *IBM systems Journal* 42.1 (2003): 5–18. Print.

Gowaty, Patricia Adair. “A Sex-Neutral Theoretical Framework for Making Strong Inferences about the Origins of Sex Roles.” *Evolution’s Empress: Darwinian Perspectives on the Nature of Women* (2013): 85. Print.

---. “Power Asymmetries between the Sexes, Mate Preferences, and Components of Fitness.” *Evolution, gender, and rape* (2003): 61–86. Print.

Gowaty, Patricia Adair, Lee C. Drickamer, and Sabine Schmid-Holmes. “Male House Mice Produce Fewer Offspring with Lower Viability and Poorer Performance When Mated with Females They Do Not Prefer.” *Animal Behaviour* 65.1 (2003): 95–103. Print.

Greene, Sheila. “V. Biological Determinism: Persisting Problems for the Psychology of Women.” *Feminism & Psychology* 14.3 (2004): 431–435. Print.

Griffiths, Paul E. “6 Evo-Devo Meets the Mind: Toward a Developmental Evolutionary Psychology.” *Integrating evolution and development: From theory to practice* (2007): 195. Print.

Gullan, Penny J., and Peter S. Cranston. *The Insects: An Outline of Entomology*. John Wiley & Sons, 2009.

Hamilton, William D. “The Genetical Evolution of Social Behaviour. I.” *Journal of theoretical biology* 7.1 (1964): 1–16. Print.

Haraway, Donna Jeanne. *Primate Visions: Gender, Race, and Nature in the World of Modern Science*. Psychology Press, 1989.

Hendry, Andrew P. et al. "Rapid Evolution of Reproductive Isolation in the Wild: Evidence from Introduced Salmon." *Science* 290.5491 (2000): 516–518. Print.

Hendry, Andrew P., and Michael T. Kinnison. "Perspective: The Pace of Modern Life: Measuring Rates of Contemporary Microevolution." *Evolution* (1999): 1637–1653. Print.

Henning, Stephen Frank. "Chemical Communication between Lycaenid Larvae (Lepidoptera: Lycaenidae) and Ants (Hymenoptera: Formicidae)." *Journal of the Entomological Society of Southern Africa* (1983).

Hoffman, Martin L. "The Contribution of Empathy to Justice and Moral Judgment." *Reaching Out: Caring, Altruism and Prosocial Behavior. MoralDevelopment: A Compendium* 7 (1994): 161–194. Print.

Hordijk, Wim. *Recognizing The Illusion Of 'Homo Economicus'.* http://www.npr.org/sections/13.7/2014/07/20/331975032/exposing-the-illusion-of-homo-economicus

Hrdy, Sarah Blaffer. "Raising Darwin's Consciousness." *Human Nature* 8.1 (1997): 1–49. Print.

Huang, Gang, Hong Mei, and Fuqing Yang. "Runtime Software Architecture Based on Reflective Middleware." *Science in China Series F: Information Sciences* 47.5 (2004): 555–576. Print.

Ickes, William. "Empathic Accuracy." *Journal of personality* 61.4 (1993): 587–610. Print.

Jax, Kurt. "Function and 'functioning' in Ecology: What Does It Mean?." *Oikos* 111.3 (2005): 641–648. Print.

Jeremy, Smith. *Our Fear of Imigrants*. http://www.psmag.com/books-and-culture/fear-immigrants-science-empathy-politics-86430

Khazan, Olga. *Multiple Lovers, Without Jealousy*. http://www.theatlantic.com/features/archive/2014/07/multiple-lovers-no-jealousy/374697/

Kruger, Daniel J. "Psychological Aspects of Adaptations for Kin Directed Altruistic Helping Behaviors." *Social Behavior and Personality: an international journal* 29.4 (2001): 323–330. Print.

Laddaga, Robert. "Active Software." *Self-Adaptive Software*. Springer, 2001. 11–26.

Laland, Kevin N., and Gillian R. Brown. “Niche Construction, Human Behavior, and the Adaptive-Lag Hypothesis.” *Evolutionary Anthropology: Issues, News, and Reviews* 15.3 (2006): 95–104. Print.

Laland, Kevin N., and Michael J. O’Brien. “Niche Construction Theory and Archaeology.” *Journal of Archaeological Method and Theory* 17.4 (2010): 303–322. Print.

Laland, Kevin N., John Odling-Smee, and Marcus W. Feldman. “Niche Construction, Biological Evolution, and Cultural Change.” *Behavioral and brain sciences* 23.01 (2000): 131–146. Print.

Laland, Kevin N., and Kim Sterelny. “Perspective: Seven Reasons (not) to Neglect Niche Construction.” *Evolution* 60.9 (2006): 1751–1762. Print.

Lawrence, E. J. et al. “Measuring Empathy: Reliability and Validity of the Empathy Quotient.” *Psychological medicine* 34.05 (2004): 911–920. Print.

Liesen, Laurette T. “Feminists, Fear Not Evolutionary Theory, but Remain Very Cautious of Evolutionary Psychology.” *Sex Roles* 64.9 (2011): 748–750. Print.

---. “Women, Behavior, and Evolution: Understanding the Debate between Feminist Evolutionists and Evolutionary Psychologists.” *Politics and the Life Sciences* 26.1 (2007): 51–70. Print.

Machery, Edouard, and H. Clark Barrett. “Essay Review: Debunking Adapting Minds*.” *Philosophy of Science* 73.2 (2006): 232–246. Print.

Machery, Edouard, and Luc Faucher, others. “Why Do We Think Racially? A Critical Journey in Culture and Evolution.” *Categorization in Cognitive Science* (2004).

Menary, Richard. *Cognitive Integration: Mind and Cognition Unbounded*. New York: Palgrave Macmillan, 2007.

Mohan, Geoffrey. *Men and Violence: Sizing up How Men Size up One Another.* Web.

Moore, A. J., P. A. Gowaty, and P. J. Moore. “Females Avoid Manipulative Males and Live Longer.” *Journal of evolutionary biology* 16.3 (2003): 523–530. Print.

Moore, Allen J., and Patricia J. Moore. “Balancing Sexual Selection through Opposing Mate Choice and Male Competition.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266.1420 (1999): 711–716. Print.

Nesse, Rudolph, and Alan Lloyd. “The Evolution of Psychodynamic Mechanisms.” *The Adapted Mind*. Ed. Jerome Barkow, Leda Cosmides, and John Tooby. Oxford University Press, 1992. 601–626. Print.

Oyama, Susan. *The Ontogeny of Information: Developmental Systems and Evolution*. Duke University Press, 2000.

Oyama, Susan, Paul E. Griffiths, and Russell D. Gray. *Cycles of Contingency: Developmental Systems and Evolution*. Mit Press, 2001.

Pinker, Steven. *How the Mind Works*. Penguin Books, 1997. Print.

---. *Reducing Violence, Increasing Empathy in the "Humanitarian Revolution."*: Audio Recording.

---. *The Better Angels of Our Nature: The Decline of Violence in History and Its Causes*. Penguin UK, 2011. Print.

---. *The Blank Slate: The Modern Denial of Human Nature*. Penguin, 2003.

Prinz, Jesse. "Against Empathy." *The Southern Journal of Philosophy* 49 (2011): 214–233. *CrossRef*. Web.

---. *The Emotional Construction of Morals*. Oxford University Press, 2007.

Prinz, Jesse J. "Is Empathy Necessary for Morality?" *Empathy: Philosophical and psychological perspectives* (2011): 211–229. Print.

Reznick, David N., and Cameron K. Ghalambor. "The Population Ecology of Contemporary Adaptations: What Empirical Studies Reveal about the Conditions That Promote Adaptive Evolution." *Genetica* 112.1 (2001): 183–198. Print.

Rushton, J. Philippe. "Genetic Similarity, Human Altruism, and Group Selection." *Behavioral and Brain sciences* 12.03 (1989): 503–518. Print.

---. "Is Altruism Innate?" *Psychological Inquiry* 2.2 (1991): 141–143. Print.

Rushton, J. Philippe, Robin JH Russell, and Pamela A. Wells. "Genetic Similarity Theory: Beyond Kin Selection." *Behavior genetics* 14.3 (1984): 179–193. Print.

Russell, Chrissie. *What's the Skinny with Triple Zero? http://www.independent.ie/life/health-wellbeing/health-features/whats-the-skinny-with-triple-zero-30485594.html*

Salehie, Mazeiar, and Ladan Tahvildari. "Self-Adaptive Software: Landscape and Research Challenges." *ACM Transactions on Autonomous and Adaptive Systems (TAAS)* 4.2 (2009): 14. Print.

Samuels, Richard. "Evolutionary Psychology and the Massive Modularity Hypothesis." *The British Journal for the Philosophy of Science* 49.4 (1998): 575–602. Print.

Scher, Steven, and Frederick Rauscher, eds. *Evolutionary Psychology: Alternative Approaches*.

Norwell, Massachusetts: Kluwer Academic Publishers, 2003. Print.

Sober, Elliott, and David Sloan Wilson. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, 1999.

Sterelny, Kim. "Thought in a Hostile World: The Evolution of Human Cognition." (2003).

Stotz, Karola C., and Paul E. Griffiths. "Dancing in the Dark: Evolutionary Psychology and the Argument from Design." *Evolutionary Psychology: Alternative Approaches. Dordrecht: Kluwer* (2002): Print.

Symons, Donald. "On the Use and Misuse of Darwinism in the Study of Human Behavior." *The adapted mind: Evolutionary psychology and the generation of culture* (1992): 137–62. Print.

Taentzer, Gabriele, Michael Goedicke, and Torsten Meyer. "Dynamic Change Management by Distributed Graph Transformation: Towards Configurable Distributed Systems." *Theory and Application of Graph Transformations*. Springer, 2000. 179–193.

Tooby, John, and Leda Cosmides. "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments." *Ethology and sociobiology* 11.4 (1990): 375–424. Print.

---. "The Past Explains the Present: Emotional Adaptations and the Structure of Ancestral Environments." *Ethology and sociobiology* 11.4 (1990): 375–424. Print.

Trivers, Robert L. "The Evolution of Reciprocal Altruism." *Quarterly review of biology* (1971): 35–57. Print.

Waal, F. B. M. de. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, Mass.: Harvard University Press, 1996. Print.

"We Need to Talk about Kevin's Lack of Empathy." *The Guardian* b 2011: Print.

Wilson, Robert A., and Andy Clark. "How to Situate Cognition: Letting Nature Take Its Course." (2009).

Historical Appendix

Researchers often take it for granted that the conceptual origin of empathy is in ancient Greek discussions of sympathy. They do so because of the similarities between Hume (1739) and Smith's (1759) treatment of sympathy and our current understanding of what empathy is. Namely, we generally take empathy to be a phenomenon whereby we place ourselves in another's shoes (taking their perspective) or of becoming aware of what it is like to be in another's situation. But while Hume and Smith's treatment of sympathy involves many processes that we now associate with empathy, they did not use the word 'empathy'.

The etymology of empathy can be traced to ancient Greek discussions of 'empatheia'. Ancient Greeks used 'empatheia' to refer to an intense individual passion or emotional experience (Depew 2005). The prefix *em-* means *into*. Accordingly, Plotinus treats empathy as the opposite of apathy (*apatheia*) —a "being into" a target, as opposed to not caring about it (*Ibid.*). For ancient Greek philosophers then, 'empatheia' meant caring about something or someone. This is very important because the first modern treatments of empathy explicitly argued for a conceptual connection to 'empatheia'.

It was Lipps who, in his works on aesthetics and psychology, argued that *empatheia* and the German word '*einfühlung*' could be understood as usefully similar (Lipps 1903a, 1903b, 1906).[55] The word 'Einfülung' had been coined in German by Robert Vischer in his doctoral dissertation in philosophy three decades prior (Vischer 1873).[56] For both Vischer and Lipps, einfülung was the *creation* by an agent of feelings and emotions in a target (Vischer 1873; Lipps 1903). It was a process of *projecting* or *feeling into* inanimate objects, plants, animals, or other humans (Lips 1903).

Not much later, Titchener (a psychologist) translated 'einfühlung' into English. He called it 'empathy':

[55] Lipps (1906) describes the connection between empatheia and einfühlung most explicitly.
[56] Depew 2005 cites Listowel (1967) as claiming that Lotze (1869) may have used the word 'einfühlung' prior to Vischer. However, Vischer et al. (1994, p 20) persuasively show that Lotze referred to a similar phenomenon, but did not coin a new term to characterize it.

> All that I have to remark now is that the various visual images, which I have referred to as possible vehicles of logical meaning, oftentimes share their task with kinaesthesis. Not only do I see gravity and modesty and pride and courtesy and stateliness, but I feel or act them in the mind's muscles. This is, I suppose, a simple case of empathy, if we may coin that term as a rendering of Einfuhlung; there is nothing curious or idiosyncratic about it; but it is a fact that must be mentioned (Titchener 1909, p 21-22).

Not long after Titchener's translation of 'einfühlung' into 'empathy', the tie between empathy and aesthetics became weaker. By the 1920's, empathy researchers were using the term to refer to process occurring mainly among humans, rather than among humans and paintings (or sculptures, willow trees, or birds).[57] For psychologists of the early twentieth century, empathy was understood to be a phenomenon whereby an agent places themselves in a target's psychological position and experiences similar feelings and emotions as the target.

[57] Why this change occurred is not well known. See Depew (2005) for some detail.

References

Depew, David. "Empathy, psychology, and aesthetics: Reflections on a repair concept." Poroi 4.1 (2005): 6.

Earl of Listowel, *Modern Aesthetics: An Historical Introduction*, New York, [Columbia] Teachers' College Press, 1967, p. 50.

Hume, David. *A treatise of human nature*. Courier Corporation, 1739/2012.

Lipps, T. Ästhetik, Teil I. Leipzig, Germany: Leopold Voss Verlag, 1903a.

Lipps, Theodor. "Empathy, inward imitation, and sense feelings." *Philosophies of Beauty: From Socrates to Robert Bridges being the Sources of Aesthetic Theory, EF Carritt (ed.)* (1903b): 252-6.

Lipps, Theodor. *Ästhetik: Die ästhetische Betrachtung und die bildende Kunst*. Voss, 1906.

Lotze, Hermann. *Geschichte der Wissenschaften in Deutschland: 7: Geschichte der Aesthetik in Deutschland*. Literarisch-artistische Anstalt der JG Cotta'schen Buchhandlung, 1868.

Smith, Adam. *The theory of moral sentiments*. Penguin, 2010.

Titchener, Edward Bradford. *Lectures on the experimental psychology of the thought-processes*. Macmillan, 1909.

Vischer, Robert. "On the optical sense of form: A contribution to aesthetics."*Empathy, form, and space: problems in German aesthetics* 1893 (1873): 89-124.

Vischer, Robert, et al. "Empathy, form, and space: problems in German aesthetics, 1873-1893." (1993).

Curriculum Vitae

Name:	O'Neal Buchanan
Post-secondary Education and Degrees:	University of East London London, United Kingdom 2005-2006 CERT
	Concordia University Montreal, Quebec, Canada 2002-2006 B.A.
	Concordia University Montreal, Quebec, Canada 2007-2008 M.A.
	The University of Western Ontario London, Ontario, Canada 2008-2015 Ph.D.
Honours and Awards:	Education for Global Competencies Bursary (HRDC) 2005
	Student Mobility Bursary (MEQ) 2005
	Mary Routledge Fellowship 2011
	Rotman Institute of Philosophy GRA 2013
Related Work Experience	Instructor, "Evil" Department of Philosophy, The University of Western Ontario 2010
	Research Assistant (Gillian Barker) "Entangled Life, Organism and Environment in the Biological and Social Sciences" 2013