

**‘FROM THE INSIDE’:
HOW TO ATTRIBUTE EMOTIONS TO OTHERS**

A thesis submitted in partial fulfilment of the
requirements for the degree of
MASTER OF ARTS
in
THE DEPARTMENT OF PHILOSOPHY
of
RHODES UNIVERSITY

By
Velislava Atanasova Mitova

December 2002

ABSTRACT

I argue that a specific version of Theory theory is necessary and sufficient for attributions and predictions of others' emotions. Theory theory is the view that we attribute and predict others' mental states on the basis of a (tacit) body of generalisations about mental states, their situational input, and behavioural output. Theory's antagonist, Simulation theory, is the view that we ascribe mental states to others by simulating - or running 'off-line' - their doxastic, emotional, and contextual situations. My argument for Theory's necessity and sufficiency develops in three stages:

First, I show that some version of Theory is *necessary* for *predictions* of *all* mental states on the basis of the ascriber's knowledge of the subject's other mental states. The linchpin of the arguments here consists of considerations from relevant similarity between the ascriber's and the subject's mental states. Simulation cannot provide criteria for such similarity, and so, I argue, predictions must advert to Theory.

Second, I develop a *sui generis* model of emotions, according to which (i) emotions' necessary *objects* and typical *causes* are concern-based construals; and (ii) emotions *qua* attitudes are (a) complex states embedded in a narrative structure, (b) characterised in terms of their object, their expressive behaviour, and their phenomenology.

Third, I show that, considering the nature of the objects of emotions, some Theory is *necessary* for emotion-predictions and -attributions. Moreover, I develop a version of Theory, based on my analysis of emotions and narrative structures, and argue that this version of Theory is both *necessary* and *sufficient* for emotion-predictions and -attributions.

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisor, Professor Marius Vermaak, for his perspicacity, support, and talent for turning my convoluted ideas into neat diagrams and lucid prose.

I am also indebted to Professor Wendy Jacobson for many fascinating discussions of the literature on which I draw, as well as for nurturing my interest in literature generally.

Thanks to Eusebius McKaiser and David Ryan for invaluable comments on an earlier draught of this dissertation; to Tom Stoneham for suggesting this topic to me; and to all my colleagues at the Philosophy Department for generously providing an incredibly happy and fruitful environment.

I acknowledge with gratitude the generous financial support from the National Research Foundation (NRF). The views expressed here are not necessarily those of the NRF.

Finally, infinite thanks to Paul for his astuteness and patience during interminable discussions of the mess of inchoate ideas which this dissertation used to be, and was going to remain without him.

CONTENTS

INTRODUCTION	1
CHAPTER 1	5
Part I: The Debate	5
Section 1: Simulation	6
Section 2: Theory	12
Part II: Heal, heuristic parsimony, and why Theory is necessary for predictions	18
Section 1: Substantive content	19
Section 2: Heuristic parsimony	21
Section 3: Theory's necessity for predictions	24
Section 4: More relevance	29
Part III: Theory and the Green Ey'd Monster	32
Section 1: Iago's apparent use of Theory	33
Section 2: Iago's incapacity for love and (Othello's kind of) jealousy	34
Section 3: Iago's 'imagination'	37
CHAPTER 2	40
Part I: Why we cannot reduce emotions to other mental states	40
Section 1: Doxastic involuntarism	41
Section 2: Belief is not sufficient for emotion	42
Section 3: Belief is not necessary for emotion	43
Part II: Emotions and their objects	47
Section 1: The Relational Schema	47
Section 2: Motivating aspects	48
Section 3: Causes as serious concern-based construals	51
Section 4: Objects as serious concern-based construals	58
Part III: Emotions as attitudes	69
Section 1: Aims	69
Section 2: Feeling towards	70
Section 3: Rationality	73
Section 4: Taking narratives seriously	77
CHAPTER 3	83
Part I: Predictions of emotions	83
Section 1: Starting in the same states	84
Section 2: Using the same mechanisms - unexperienced emotions	86
Part II: Attributions of emotions	91
Section 1: Theory's necessity for emotion-attributions	91
Section 2: Theory's sufficiency for emotion-attributions	97
Section 3: Theory's sufficiency for emotion-predictions	102
CODA: Under the carpet	105
BIBLIOGRAPHY	113

INTRODUCTION

A central debate in current philosophy of mind revolves around “the basis question”: “what is the basis for our ability to engage in folk psychological practice?”¹. The main antagonists in the debate are Simulation theory and Theory theory. Philosophers focus on three levels of folk practice, some or all of which interest different participants in the debate. The first level concerns our understanding of the *nature* of mental states (“the metaphysical question”²). The second level concerns our possession and manipulation of mental state *concepts* (“the concept mastery question”³). The third level is about how we *attribute, predict* and *explain* others’ mental states (“the explanatory question about normal adult folk psychological practice”⁴). The last question will interest me in this dissertation. In particular, I concentrate on prediction and attribution. *Simulation*⁵ is the view that we ascribe mental states to others by simulating their doxastic, emotional, and contextual situations. *Theory*⁶ maintains that our attributions and predictions are based on a tacit theory we possess, an organised body of knowledge wrought of quasi-nomological generalisations about mental states, and their situational input and their behavioural output.

Jane Heal, in ‘Understanding Other Minds from the Inside’, has claimed that Simulation is the only approach that allows for understanding others first-personally, that is, in a way which “find[s] the other person’s life intelligible ‘from the inside’”⁷. By contrast, Theory theory, she avers, “does not seem at all hospitable to the ‘from the inside’ idea”⁸. This dissertation takes up Heal’s gauntlet, by focusing on our most personal attributions - emotion-attributions. I argue that the hospitality which Simulation seems to proffer to the ‘from the inside idea’ *à propos* emotion-attributions and predictions is spurious. My central claim is that *Theory is necessary and sufficient for both emotion-predictions and emotion-attributions*.

My argument unfolds in three stages. First, I show Theory necessary for predictions of *all* mental states, based on the ascriber’s knowledge of the subject’s other mental states (Chapter 1). Second, I develop a *sui generis* view of emotions according to which: (a) their objects are *concern-based construals*; and (b) attitudinally, they are characterised in terms of

¹ Davies and Stone (1998: 55).

² Stone and Davies (1996: 119).

³ *ibid.*

⁴ Stone and Davies (1996: 120).

⁵ Throughout this dissertation ‘Simulation’ stands for the Simulation *view*; while ‘simulation’ denotes the particular ascription/ prediction *procedure* Simulation envisages.

⁶ I use ‘Theory’ as shorthand for the Theory *view* throughout; while ‘theory’ refers to any particular metarepresentational knowledge we need to invoke for individual ascriptions.

⁷ Davies’s and Stone’s gloss (1998: 81).

⁸ Heal, (1998: 84).

their *phenomenology*, *ensuing behaviour*, and *narrative structure*. (Chapter 2). Third, I show that Theory, properly understood, is necessary and sufficient for predictions and attributions of emotions precisely because of the first-personal nature of concern-based construals (Chapter 3).

1. The arguments

Chapter 1 develops in three stages. First, I map out the debate between Theory and Simulation. I settle on the most inclusive versions of both positions, in order for the choice between the two to be exhaustive. In the process, I raise some general doubts about Simulation's capacity for ensuring that the mental states which the ascriber simulates are *relevantly similar* to the subject's mental states (Part I). Second, I explore Heal's arguments for Theory's heuristic promiscuity and Simulation's heuristic parsimony, which turn on the consideration that substantive content is necessary for predictions. I argue that it is precisely considerations from substantive content that render Theory (a) more parsimonious than Simulation, (b) necessary for predictions, and (c) more hospitable to the from-the-inside idea (Part II). Finally, I consider Iago's predictions of Othello's emotions, and argue that Iago not only *appears* to be using Theory but is also incapable of Simulating Othello, and so *must* be using Theory (Part III).

Chapter 2 grapples with the question of what sort of mental states emotions are. I argue that emotions are *sui generis*. The most promising reduction-candidate (either solo or in combination with desire) is belief. I argue that belief is neither necessary nor sufficient for emotions. It follows that emotions cannot be reduced to beliefs, judgements, or any combination of belief/ judgement and other mental states (Part I). The positive account of emotions has two stages. First, I consider emotions' *objects* as captured by de Sousa's Relational Schema. I argue that the Schema needs to be refined by casting the emotion object's various aspects in terms of Roberts's notion of 'concern-based construals'. Such construals emerge as the typical causes and necessary objects of emotions (Part II). The second stage of the account focuses on emotions *qua attitudes*. The only attitudinal aspect that the Relational Schema canvasses is an emotion's typical ensuing behaviour. To this I add a new element - the characteristic 'feeling towards' the emotion's object (which includes a certain *seriousness* towards the construal as well as characteristic *phenomenology*). Finally, each emotion has a paradigmatic narrative structure which serves as a palimpsest for judgements of appropriateness (Part III).

Chapter 3 develops in two stages. First, I show that if Simulation is correct, then knowledge of the full causal history of a particular emotion is *necessary* for *predicting* that emotion (Part I). I then argue that knowledge (a) of the Relational Schema of a particular

emotion, (b) of paradigmatic relations between the emotion type and character types, and (c) of the subject's character, is *sufficient* for emotion-attributions and *predictions*. Since possession of (a) to (c) is possession of a Theory, Theory emerges as sufficient for attribution and predictions.

2. Methodology

One aspect of my otherwise analytic methodology may seem puzzling, so let me say something about it here. I make use of examples from literature throughout the dissertation. I do so rather shamelessly, with no attempt at furnishing the method with some theoretical underpinnings. This would be fine, the envisaged worry goes, if I were using literature merely for cosmetic purposes, but I actually use it to generate *substantive* arguments.

Here is my excuse. Thought-experiments and examples in general are a legitimate tool in philosophy, especially in an area where we are analysing actual psychological practices. My problem is that such examples always leave a taste of the prefabricated, and are distilled *ad nauseam* in response to objections. Such disillusionment with examples can be allayed in two ways. One could use real data from cognitive science and join the current *tu quoque* between Simulation and Theory, based on the data. Alternatively, one can draw on literature which abounds in paradigmatic situations of emotions, and of their attributions and predictions. I choose the latter.

Of course, there is something ironic about claiming that *fiction* can provide us with *real* examples of emotions and attributions. The irony is apparent only, I contend, for three reasons. First, only fiction offers situations and people, described *richly* enough not to be prone to philosophical gerrymandering. Second, we all have *access* to the full description in a way that militates against *ad hoc* elaborations. Third, and most important, good fiction provides *accurate* descriptions of situations and people. The accuracy is evinced in the longevity of certain works. I have chosen two masters for the purpose - Shakespeare and Dickens. But the underlying intuition is *not* that because they are *masters* we should heed their vision. This would be, I appreciate, an unwarranted appeal to vintage and authority. Rather, we are moved by them and have hailed them as masters because they have captured something significant about human experience. And it is because they have faithfully crystallised something about human experience that philosophers should heed them. As far as I am concerned, this is all the 'theoretical' underpinning that the use of literature requires.

3. Structure

The biggest unit in the dissertation is a chapter. Next come parts. A part is divided into entitled sections. *Sub*-sections (also entitled) are labelled 1., 2., etc. When a further division is necessary, I label *sub*-sub-sections a., b., etc. When *sub*-sub-sub-sections (if I may be forgiven) are necessary, I label them (i), (ii), etc. When I need to distinguish claims which do not merit a whole section, I simply flag them with (a), (b), etc.

CHAPTER 1

Introduction

In the present chapter I argue that, on the broadest possible understanding of both Theory and Simulation, Theory is necessary for *predictions* of *all* mental states on the basis of the predictor's knowledge of the subject's other mental states. Chapter 3 will complete the argument by showing that a *specific* version of Theory is necessary and sufficient for *emotion-predictions* and attributions.

Chapter 1 unfolds as follows. In Part I, I map out the Simulation-Theory debate by describing several versions of each. I settle on the broadest versions of both, since it is only then that the choice between the two becomes exhaustive. In Part II, I focus on Jane Heal's account of Simulation and argue:

- (a) against her claim that Simulation is more parsimonious than Theory;
- (b) that both her relatively minimal account and a more elaborate one will need to advert to Theory at some stage of the envisaged simulation procedure.

In Part III, I consider Iago and argue that he predicts Othello's emotions on the basis of a theory.

Part I: The debate

Introduction

The aim of this part is to describe the Simulation-Theory debate. In section 1 I elucidate the notion of simulation posited by Simulation theory and trace some of its implications. I then describe the two mainstream versions of Simulation: those relying on an analogical inference from the ascriber to the simulated subject (Goldman), and those which envisage imaginative "transformation" rather than inference (Gordon). I suggest that, *prima facie*, neither appears to offer much hospitality to the 'from the inside idea', the idea that

in giving a psychological explanation we render the thought or behaviour of the other *intelligible*, we exhibit them as having some *point*, some *reasons* to be cited in their defence (Heal, 1995a: 52, my italics).

In section 2 I consider the three dominant versions of Theory: first, a version based on an analogy with linguistic theory; second, a version which conceives the folk's theory as a quasi-scientific theory; finally, the broadest version, which credits the folk with nothing more specific than some sort of body of psychological knowledge.

Section 1: Simulation

1. *The account*

This is how Martin Davies and Tony Stone, the most prominent chroniclers of the debate, describe Simulation's vision of predictions:

[T]he simulation theorist sees engagement in the folk psychological practice as *re-enactment*...The simulation strategy involves using imagination to cantilever out from our own theoretical and practical reasoning - leading to judgements and decisions - to an understanding of the beliefs and actions of another person (1996:128, italics in original).⁹

The idea is that we are naturally endowed¹⁰ with mechanisms for (a) practical reasoning, (b) theoretical reasoning, and (c) emotion-response¹¹, which we all share, and employ (a) when we are making decisions about our own lives; (b) when we engage in thought about the non-actual ('what would I *think* if...' and 'what would I *do* if...'); and (c) when we emote¹². The simulationist claim is that when we think about others - what they would think, do, or feel - we use these mechanisms to make predictions. We feed the mechanism with 'pretend' input (beliefs, desires and emotions), let the mechanism run 'off-line'¹³ and wait for it to churn out an output. The output then is the other's predicted belief, desire, emotion, or action. This is the sub-personal account. The personal account replaces talk of feeding a mechanism with talk of imaginative identification with the other. But here, too, it is understood that the prediction work is done by shared mechanisms (e.g. Heal, 1998).

2. *Assumptions*

Simulation's claim rests on some assumptions of similarity.

a. *Shared mechanisms*

The first, and least problematic, assumption is that we have *shared* mechanisms for practical and theoretical reasoning. The assumption seems unproblematic because, after all, we do share

⁹ 'Emote' is the verb I use throughout this dissertation. I appreciate that it lacks grace, but I need just such a neutral term, unique to emotions (as opposed to something like 'feel' or 'experience emotion').

¹⁰ Most simulationists view this endowment as only partly innate, and mostly as developed.

¹¹ Nichols *et al.* (1996) are the only writers who distinguish amongst these three mechanisms and, correctly point out that simulationists generally and mistakenly maintain that all kinds of simulation-procedures rely on our practical reasoning mechanism. Since I am trying to offer the most plausible description of Simulation here, I incorporate their insight. They articulate it thus:

Given a boxology (a functional architecture), *each* component can be viewed as a possible engine of simulation. In principle *any* component can be taken off line (detached from its usual function) and be used to perform some other function (1996: 41, italics in original).

¹² 'Emote' is the verb I use throughout this dissertation. I appreciate that it lacks grace, but I need (especially in Chapter 2) just such a neutral term, unique to emotions (as opposed to something like 'feel' or 'experience emotion').

¹³ Gordon (1995a, 1995b, 1996, 1998).

a common description of these mechanisms, whose locus are principles of rationality. Indeed, the most plausible version of Simulation is one which confines its claim to predictions in the “domain of reason” (Davies and Stone 1998: 81, Heal 1998). But it should be understood that, according to Simulation, principles of rationality merely *describe* the mechanisms, they do not *qua theoretical entities* figure either in our own thinking about ourselves or in the aetiology of our predictions.

b. *‘Pretend’ versus actual input*

The second assumption Simulation makes is that feeding these mechanisms with *pretend* input is sufficiently similar to feeding them with *actual* input, that is, that the *mechanisms* do the same job when fed with pretend input as they do when fed with actual input. Again, this is not very problematic if we concede the claim that when I think about the non-actual in my own case, I use these mechanisms and feed them pretend input successfully. Of course, one possible problem for this assumption would be our embodiment: our own mechanisms are often affected by “non-rational influences on thinking” (Heal, 1995a: 48), like, say alcohol. So, the objection might go, there is no way of either eliminating *my* non-rational influences when simulating another not so influenced, or of letting the non-rational influences of the other be relevant to my simulation of him without some theory. At this stage, however, I am more interested in seeing whether Simulation can work in the ideal case, so let us ignore “non-rational influences”.

c. *The input's content*

The third, and most problematic, assumption is that *my* pretend belief which I feed into my mechanism in a prediction of A is sufficiently like A's actual belief. There are two issues here: attitudinal and content-related.

(i) The attitudinal issue amounts to the assumption that *qua attitude* there is no difference between a pretend belief and an actual belief. I think this collapses into the second assumption above: since we are interested in predictions, a belief *qua attitude* will be defined in terms of its functional role, but that in turn means that a pretend belief is the same as an actual belief just in case it plays the same role in the practical and theoretical reasoning mechanisms. So the attitudinal aspect of the assumption should not worry us more than the second assumption itself.

(ii) The content assumption is the most problematic. I will keep returning to it throughout the dissertation. For the moment, it should be noted that we will have trouble making the assumption stick if:

(a) we notice that substantive content is necessary for predictions (a corollary of Heal's 'from the inside idea');

(b) and accept the holism of the mental (which Simulationists do and in fact argue *from*, e.g. Heal).

The reason (a) and (b) will be a problem is that for Simulation to capture a prediction of *another's* mental state, both the input and the output 'pretend' states must have the same content as the subject's. But if we accept that the mental is holistic, we accept, in effect, that the feeding of a single 'pretend' belief into the simulation mechanism, involves a plethora of concepts and related beliefs. This means that for the simulation to be a prediction of another's mental state, my 'pretend' input state must be identical in content to the other's state, which in turn implies that my relevant satellite concepts and beliefs must be identical.

3. *Versions of simulation*

In this context, there are two kinds of Simulation positions.

a. *Goldman*

The first envisages particular simulations as involving a special kind of analogical inference, of the following form: 'If he is psychologically like me, he must be in mental state M; he is psychologically like me; therefore, he is in mental state M (Goldman, 1995a: 92).

Goldman, who is considered the most prominent proponent of the inference view, denies that this is his position. In response to standard objections against the analogical inference, he writes:

The best line of reply is...to deny that interpreters must believe the second premise. Many beliefs are formed by mechanisms, or routines, that are built into the cognitive architecture. Although these mechanisms might be described in terms of 'rules' or 'principles', it would be misleading to say that the cognizers believe those rules or principles. For example, it is plausible to say that people form perceptual beliefs...in accord with Gestalt rules, but implausible to say that they literally believe those rules (1995a: 93).

And the idea is that the belief that the outcome of a simulation is a prediction of another, is informed by the 'rule' that he is similar to me, in the same way as perceptual beliefs are informed by Gestalt principles. The passage can be read as Goldman's repudiation of the analogical *inference*. It is clear, however, that although 'the interpreter'¹⁴ does not make the inference in particular simulations, he must either be credited with it, or not be considered an

¹⁴ Goldman's 'interpreter' denotes a broader notion than my predictor or ascriber: it includes also a person trying to *explain* and *understand* another's behaviour and mental states.

interpreter of another at all. The passage reveals an inherent tension within this strand of Simulation, generated by, on the one hand, the need to substantiate the content-assumption, and on the other hand, wanting to downplay it. The suggestion is that the assumption is somehow supported by the actual and subpersonal realisation of the analogical inference. This, of course, renders the status of the output state dubious in the sense that it is unclear how it can count as a prediction of *another's* mental state. The only epistemic model, as Goldman realises, that can call this a prediction of another is reliabilism (since only it allows for knowledge which need not be based on justification). So, this version of Simulation commits its friends to reliabilism. Such a commitment would be an embarrassment to a herald of the from-the-inside idea which turns on reasons and justification.

b. *Gordon*

The second Simulation position, developed by Gordon, proposes to avoid the problems of the analogical inference by claiming that simulation does not involve an “inference from me to you” (1995d) but rather a “transformation”. The envisaged simulation is still run on my own “motivational and emotional resources, and [my] own capacity for practical reasoning” (1996: 11). But I do not then translate the output into a prediction, that is, I do not pick up the output and think ‘This is what I would have done/ thought/ felt if I had these beliefs, and so it is what A will do/ think/ feel’. Rather, from the beginning the ‘I’ in my practical or theoretical reasoning stands for A, and so the output naturally and automatically is about A.

I will look at an example from *A Midsummer-night's Dream* that Gordon offers. Although it is about predicting *action*, and I am concerned with predictions of mental states, it is useful for elucidating Gordon's notion of transformation. Hermia awakes in the forest and discovers that Lysander is missing. Her immediate response is to blame Demetrius for murdering Lysander. How would she go about retrodicting¹⁵ that?

she would transport herself in imagination into [Demetrius's] situation to the extent to which it seemed, to a first approximation, relevantly different from her own; but not strictly transporting *herself*, Hermia, but rather a self *transformed*, in so far as seemed necessary, into someone who would behave as she had known Demetrius to behave (1996: 12, *italics in original*).

¹⁵ In this case an account of retrodiction would be the same as an account of prediction, since in both cases the simulator must put himself in an initial situation, and does not know the final outcome. That is, it is irrelevant whether Hermia attempts to predict Demetrius's murder of Lysander when she is still at Theseus's palace or in the forest, since in neither case does she know that Lysander is murdered. An account of standard retrodiction would be different, and closer in nature to an account of *explanation* of others' mental states/ behaviour.

I will have more to say about this adjustment for relevant differences, in both Part II below and Chapter 3. For the moment two central assumptions underpinning Gordon's notion of transformation need to be noticed. Gordon denies that either of two sets of concepts is necessary for a prediction:

- (i) concepts of the mental states which figure in the causal story of the state predicted;
- (ii) the concept of the actual mental state predicted (in the case of predictions of mental states, as opposed to the behaviour-prediction in Hermia's case).

(i) One of Demetrius's mental states that is causally relevant for (Hermia's prediction of) Demetrius's killing Lysander is Demetrius's love. Gordon argues that Hermia need not have the concept of love in order to simulate the envisaged prediction:

Decisions made in the role of another agent can be influenced by unrecognised emotions, *including* any emotions picked up from the person whose actions are being predicted...if Hermia were predicting Demetrius's actions by simulating Demetrius, then she would not have to *categorise* her second-order emotion at all - she would only have to *use* it, that is allow it to do its usual work of influencing behaviour (1996: 14, italics in original).

The passage points to the role that emotion *contagion* plays in Gordon's account. The idea is that we sometimes 'pick up' people's emotions just as we pick up their colds or yawns. It is clear that I do not need to have the concept of the emotion (or cold, or yawn) in order to have the emotion; and once I have it, it influences my own decisions and reasoning just as an ordinarily acquired emotion would. So far so good. But, one may want to ask two questions. First, can I just pick up other mental states, like beliefs, in this manner?

Second, is there not something arbitrary about a prediction whose underpinning mechanism involves contagion? For example, if Demetrius resembled Hamlet, one may have picked up from him not love but depression. Now this depression, although (suppose) the symptom of unrequited love, would not have the same function if picked up from Hamlet and plugged in Hermia's simulation mechanism for predicting Hamlet's actions. In Hamlet, in other words, *two* mental states would play their causal and motivational role: depression and love. But Hermia would only pick up depression through contagion and that is the only mental state which will play a causal role in *her* mechanism. The objection, then, is that for the envisaged simulation to be a genuine simulation of another, contagion cannot be sufficient. What would make it sufficient is some *inference* to the effect that depression in Hamlet betokens unrequited love and so love should play a role in my simulation. But this means that Hermia must

conceptualise (“categorise”) the causally relevant mental states after all. I will return to this point in Part II of this chapter (section 3).

(ii) The second denial that Gordon makes is that the concept of the predicted state is necessary for a prediction. The denial is underpinned by his account of self-knowledge: the claim is that in our own case we do not need the concept of belief in order to know that we believe *p*. All we need to do, in order to discover whether we believe *p* is an “ascent routine”, that is, ask ourselves the question “Is *p* true?” and if we (sincerely) say yes, then we believe *p*. He does concede that the ascent routine does not lead to “*genuine, comprehending* ascriptions of belief” (1996: 15, italics in original), but he thinks that this is precisely where Simulation plays a vital role:

the very process of embedding ascent routines in simulations remedies this deficiency. It gives sense to the idea of a mental location...the notion of something's being a fact *to* a particular individual (1996: 18, italics in original).

The idea is that this is how we develop the concept of belief: as children, we run simulations of others, and by the imaginative transformation involved (running off-line A's states as mine where ‘I’ is A), we learn the difference between what are facts *to* me and what are facts *to* other people. And presumably, the obvious threat of circularity in this bootstrap (as Gordon calls it) would be averted by noticing that the full-blooded concept of a mental state is not needed in a simulation. All we need to grasp is this notion of a fact *to* another (which at some point becomes the concept of belief).

I am uncertain how this model would work for predicting *desires* or *emotions*. While both involve some representations of the world (and so some facts *to* the agent) such representations are certainly not exhaustive of either mental state. Of course, we could go for full-blown reductionism (desires and emotions just are beliefs/ judgements), but this seems a dear price to pay for salvaging Simulation. Alternatively, we could follow Nichols *et al.* by postulating separate mechanisms for the predictions of different kinds of mental states (cf. fn. 11, p. 6 above), and different conditions for each. This would be fine, but it would lead to a discontinuous and *ad hoc* account of self-knowledge.

These, then, are some of the elements and assumptions of Gordon's model. Does it assuage my worries about content, generated by Goldman's account? Patently, no. Part (i) of Gordon's account seemed promising - we do not need a substantive notion of similarity because we do not need, according to Gordon, to conceptualise the causal story of other mental states involved in the mechanism leading to the output mental state. Two points make this account unsatisfactory. First, I disputed that Gordon is right about this. Second, my initial worry was

about the substantive *content* of the *input* mental state. Gordon's account (if it were plausible) would resolve problems of similarity to do with the *mechanism* involved and with the *output* state. As I said earlier, however, what poses a problem for someone who wants to defend Simulation *and* endorses the holism of the mental, is a substantive notion of the concepts involved in the propositional content of the *input* mental state. Gordon has not even mentioned that. But it is clear that it is precisely the substantive content of a mental state which would determine whether a simulator has *genuinely* "transformed" himself into the other. I will have more to say about these two levels of relevant similarity in Part II (sections 2 and 3).

This then is the picture of Simulation, and some of its problems. It seems that both strands of the position are poorly equipped to handle substantive content and so are *prima facie* inimical to the 'from the inside idea'. Before I turn to a more detailed examination of this animosity (Part II), I explore the Theory position and see what it has to offer on this score.

Section 2 - Theory

The central tenet of Theory is that when predicting others, we *use* a (perhaps *tacit*) body of information about how our psyches function. So, for example, if I knew that A knows that if it rains then his garden will be wet, and that A believes that it is now raining, I use *the knowledge that modus ponens governs our reasoning*, in order to predict that A believes that his garden is wet. If I knew, in addition, that A was planning to celebrate his birthday outdoors and had bought meat for the barbecue, and it is the first time he would bring all his friends together, then I can predict that he will be pretty upset when it rains. I would do this by some logic (if rain, then no barbecue), and by some psychological generalisations (it is vital for the self-conception of people who have different groups of friends and crave a non-schizophrenic life, that they bring their friends together; people whose self-conception projects are frustrated, tend to be upset, etc.). Of course, I neither have to have an *articulated system* of these precepts, nor need to make *conscious use* of them. Indeed, postulating a fully articulated system and conscious use would be phenomenologically incorrect. I will return to this point presently (p.13).

According to Davies and Stone¹⁶ there are three ways of fleshing out the Theory position: by analogy with linguistic theory; by analogy with scientific theory; and by crediting

¹⁶ Much of the expository part of this section is borrowed from Davies and Stone (1995a; 1995b, 1996, 1998). I presume to do so, because a description of the debate is not the aim of this dissertation. Rather, I intend to contribute to the debate. In light of this it makes sense to use the most general and accurate possible description of it, and then offer my own work.

the folk with a very general body of information about how mental states interrelate with each other and with the environment. The assumption underlying all three alternatives is the “dominant explanatory strategy”. Stich and Nichols describe the strategy as follows:

[it] proceeds by positing an internally represented ‘knowledge structure’ - typically a body of rules or principles or propositions - which serves to guide the execution of the capacity to be explained. These rules or principles or propositions are often described as the agent’s ‘theory’ of the domain in question. In some cases, the theory may be partly accessible to consciousness; the agent, can tell us some of the rules or principles he is using. More often, however, the agent has no conscious access to the knowledge guiding his behaviour. The theory is ‘tacit’...or ‘sub-doxastic’ (Stich and Nichols, 1995a: 123)¹⁷.

The three versions agree that some body of information is *causally active* in - as opposed to merely *descriptive* of - our capacity for predictions and explanations. They envisage differently, however, the nature of this body of information.

1. *The linguistic analogy*

The central claim of this version of Theory is that the envisaged body of knowledge is governed by rules similar to the ones governing linguistic knowledge. When the linguist tries to explain our “ability to produce and understand an indefinite number of sentences in [our] native language”, he proceeds as follows:

the explanation provided by the linguist begins with the postulation of a grammar. The grammar...is such that, were it to be known and deployed by the language user, it would result in the speaker being able to produce and understand just the sentences that she does in fact produce and understand...The application of the ‘dominant explanatory strategy’ comes when the linguist claims that a person who knows a language does so *in virtue of* being in possession of the body of knowledge expressed by the grammar (Davies and Stone, 1995a: 8, my italics).

So the idea is that it is in virtue of (tacit) rules and structures, analogous to grammatical ones, that we explain and predict others.

The first problem both for the linguistic explanatory strategy and for the folk one modeled on it, is the tacitness involved. That is, we need to describe our folk grammar/ theory as to some extent subconsciously *possessed* and *used*. Such talk of tacitness seems to render Theory unfalsifiable. The simulationist should not get too excited about this, though, because Simulation is prone to the same charge. Many simulationists (e.g. Goldman, 1995a: 88) acknowledge that we are equally phenomenologically unaware of simulating others in

¹⁷ The passage is also cited by Davies and Stone (1995a: 8 & 1996:124), but in order to elucidate the *linguistic analogy* (which I think is a mistake).

predictions. This makes sense: if we were aware of simulating or using a theory, the debate would not have been raging for over ten years. So Theory and Simulation are on the same footing here.

The second problem for the analogy, as noticed by Davies and Stone, is that a native language speaker can be perfectly competent without ever having a notion of the grammar in question, while the folk seem to possess a lot of the relevant concepts envisaged by Theory. Tacitness, if we recall Stich's and Nichols's description of the 'dominant explanatory strategy', was a matter of how accessible to consciousness the underlying Theory-principles are. And it is clear that if a native speaker can be competent without knowing the principles, these principles can be (and often are) totally inaccessible to consciousness. By contrast, the body of folk principles seems much more accessible, as evinced by our often citing such principles in explanations. This point, developed by Botterill (1996: 114) in *defence* of Theory, counts *against* a conception of Theory modeled on the linguistic analogy.

2. *The science analogy*

The most prominent trait of this version of Theory is that it characterises the envisaged body of psychological information by analogy with the *deductive-nomological* principles organising various scientific programmes. Folk explanation of others, then, becomes explanation in terms of "subsumption under generalisations" (1996). Davies and Stone cite Lewis (1972) as the chief proponent of this view. He describes the folk's theory as follows:

Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli and motor response...Include only platitudes which are common knowledge among us - everyone knows them, everyone knows that everyone else knows them, and so on (cited in Stone and Davies, 1996: 121-2).

The ensuing body of information, then, is a combination of "deductive-nomological explanations" and their relations to "how things generally tend to happen" (Stone and Davies, 1996: 122).

There are several problems with this model of Theory. The first is that we do not have a nearly so well developed and precise *description* of folk theory as we do of scientific theories. The second is that most of the time we do not make conscious *use* of these principles when predicting others. These two problems lead to the postulation of a tacit body of knowledge and a tacit use of it, which has its own problems as described above. The good news in this case, however, is that Botterill's suggestion that these principles are quite accessible to us, as evinced by our citing them in explanations, favours this version of Theory.

The third problem with Theory conceived on the scientific analogy is that it enlivens Heal's charge of hostility to the 'from the inside idea'. As Stone and Davies suggest, scientific theories, and so the analogical version of folk Theory, "must contain generalisations of the right kind - objective, counterfactual supporting - [in order] to figure in subsumptive explanations" (1996: 81). This amounts to the quasi-scientific model of Theory treating agents as "complex objects", as Heal complains. But, she correctly points out, folk psychology does not actually treat people like this. The 'from the inside idea', recall, was the idea that:

in giving a psychological explanation we render the thought or behaviour of the other *intelligible*, we exhibit them as having some *point*, some *reasons* to be cited in their defence (1995a: 52, my italics).

Heal's problem, then, is that scientific explanations *essentially* are *objective* and *aetiological*. Psychological explanations are, by contrast, *essentially personal* and *reason-providing*. This version of Theory, then, is implausible if it takes the science-analogy too far.

3. *Theory as an unspecified body of psychological information*

The best response to all these problems is to remain silent about the nature of the envisaged body of psychological knowledge which the folk possess. This may seem like an evasion of the issues at stake, but three considerations should dispel such doubts. First, as Nichols and Stich point out (1995a: 133), the more liberal our notion of Theory is, the more exhaustive the choice between Theory and Simulation becomes. That is, on a broad version of both, arguments against the one are arguments for the other. Second, only once the choice is made is there any point in fleshing out either account. Finally, as should be clear from the preceding discussion, we can offer some general principles which constitute the folk's theory.

Of course, we need to be cautious here. One danger is that we may end up with such an inclusive notion of Theory that it no longer defines any interesting position distinct from Simulation. For example, many hybrid theorists have suggested that the folk's capacity for predictions is grounded in *theory-driven* as opposed to *process-driven* simulation (e.g. J. Perner (1996), Botterill (1996), P. Carruthers (1996)). Since the claim of this chapter is only that *some* Theory will be *necessary* for predictions, I do not need to be too particular at this stage about Theory. Still, it would be good to put some flesh on these bones before I go on.

Here is the suggestion, then. First, Theory can be *partly* modeled on the scientific analogy, more precisely partly modeled on an analogy with the special sciences whose laws are hedged. This accounts for the quasi-deductive-nomological character of the explanations we offer when we *causally* explain each other's mental states and behaviour. But this cannot be the whole story, because we also explain each other in terms of reasons. So we need some

provision for them: part of our body of psychological knowledge, I submit, consists of precepts about the conceptual links (a) among mental states; and (b) among mental states, situational input, and ensuing actions. This will be sufficient to satisfy Heal's requirement for explanations in terms of reasons, while not committing her mistake of thinking that we *only* offer explanations in terms of reasons (consider: 'Oh, don't be angry at him, he has had too much to drink').

With regards to the second part of the proposal, Chapter 2 will be devoted to spelling out precisely what these links are in the case of emotions. In Chapter 3 I will argue that the proposal is necessary and sufficient for *emotion*-attributions and -predictions.

The proposal can be refined by consideration of Botterill's position. He suggests that "there is a core to folk psychology which functions in much the way that a hard core informs a Lakatosian research programme" (1996: 112). The three main principles which Botterill places at the core are:

[Action Principle] An agent will act in such a way as to satisfy, or at least increase the likelihood of satisfaction of, his/her current strongest desire in the light of his/ her beliefs.

[Perception Principle] When an agent A attends to a situation S in a given way, and p is a fact about S perceptually salient in that way, then A acquires the belief that p.

[Inference Principle] When an agent A acquires the belief that p and a rational thinker ought to infer q from the conjunction of p with other beliefs that A has, A comes to believe that q. (1996: 115-116).

Notice that the principles simultaneously capture appropriate *causal* connections and allow for explanations in terms of *reasons*. This is consonant with, and a refinement of, my proposal. Where Botterill and I diverge is that he places at the periphery (to pursue the Lakatos simile) what he calls "folksy psychology" (a body of apothegms, rules of thumb, and generally old-wifish principles), and insists that their implementation in predictions needs some simulation. I would urge, by contrast, that those constitute precisely the main body of the non-tacit part of Theory. Adages like 'Cruel is as cruel does' are both fundamental to our conception of others and are underpinned by the more rigorous conceptual links between action and mental states

captured in Botterill's Action Principle¹⁸. To pursue another analogy with science (this time developed by Stone and Davies, 1996: 126), these apothegms resemble our theory about the continuity and solidity of the objects around us. Though its rigorous scientific descriptions are quite different and not accessible to many, we do just fine interacting with objects on the basis of this less fine-grained folk theory. Indeed, it is this theory which does the causal job in predictions about physical objects.

Conclusion

I have outlined in this part of the chapter, the debate between Simulation and Theory. I have claimed that only the broadest interpretation of Theory allows for an exhaustive choice between Simulation and Theory. The debate, then is between the following two broad positions:

Theory is the view that a body of psychological knowledge figures causally in our predictions (and explanations) of others;

Simulation is the view that such a body can *describe* our prediction-mechanisms, but is not causally operative in them. Instead, it is the subpersonal mechanisms so described that do the causal work.

Since the choice is exhaustive, arguments against Simulation are arguments for Theory and arguments against Theory are arguments for Simulation. Part II of this chapter grapples with Heal's claim that Simulation is more heuristically parsimonious than Theory (a standard argument against Theory); and shows that not only does Theory seem more parsimonious, but it is also necessary for predictions.

¹⁸ Marius Vermaak has pointed out correctly that my position is affiliated with Elster's notion of mechanisms (1999). There are three reasons why I do not capitalise on this similarity. First, I am not confident that I understand fully Elster's notion of mechanism (especially his distinction between A-mechanisms and B-mechanisms). Second, talk of mechanisms sounds too unnervingly similar to Simulation-speak. So if I adopted it, the distinction between my position and Simulation would be diluted. Recall that Simulation does not deny that the mechanisms in question can be *described* in Theoretical terms, though Theory is *causally* impotent. Championing use of mechanisms seems to entail (by the definition of mechanism) Theory's causal impotence. Finally, one of Elster's arguments for mechanisms turns on the consideration that the folk has incompatible adages (cf. Part II, section 2). For him this is an advantage. If I, however, claim that all adages (including conflicting ones) are symptomatic of the *correct* (as opposed to actual) theory that Theory envisages, I would resurrect the problems I will impute to Heal of assuming that two agents would arrive at the same solution to a problem or would arrive at it in the same way (cf. Part II, section 2). Of course, I am not proposing that we ignore conflicting adages, just that we explain their existence by either denying that one of the incompatible pair is correct, or by claiming that different adages capture insights about different character types (cf. Chapter 3, Part II).

PART II: Heal, heuristic parsimony, and why theory is necessary for predictions

Introduction

In her 'How to think about thinking' Jane Heal proposes that Simulation and Theory theory both contribute important insights to a correct conception of mental-state attribution. She does this by arguing that the two approaches differ in their applicability across three aspects of our intercourse with others' mental states: (a) *attribution* and *prediction* of a particular thought on the basis of the ascriber's prior knowledge of the subject's *other mental states*; (b) *understanding* of what is involved in judgements about others' thoughts in general; and (c) *attribution* and *prediction* of a particular thought on the basis of knowledge of the subject's *behaviour* and *circumstances*. She argues that Simulation is the most apposite approach to the first case and may play some explanatory role in the latter two cases¹⁹.

In this part I focus on the first case, particularly on *predictions*. I argue that even in this ideal case - when we know the subject's other mental states - Simulation does not display heuristic clout superior to Theory's. I do this by considering Heal's arguments and showing that they do not sustain her anticipated conclusions but, on the contrary, favour Theory.

One aspect of my methodology may seem puzzling, so let me say something about it here. Since Part II explores exclusively Heal's account of Simulation as developed in four seminal papers, it may be doubted that I achieve my intended aim - to subvert Simulation's plausibility *in general*. There are several reasons for this narrow emphasis, all stemming from the dual purpose of Chapter 1, which is to *introduce* the debate and confer *prima facie* plausibility on Theory's necessity for predictions. Much of the debate has a *tu quoque* air about it, which, I submit, can only be purified by engaging with actual arguments and preferably good ones. So Heal is the appropriate antagonist, firstly, because all the central issues in the debate and problems with Simulation emerge from her account. This should be obvious from the discussion of the two main versions of Simulation in Part I. Second, most of the positive work for Theory's necessity will be accomplished in Chapter 3 which is much more general, and so it would be neat to have a specific view to refer back to and to serve as a

¹⁹ Because of that, she claims, "simulation and theory should not be seen as mutually exclusive rivals" (1995b: 34). I have three reasons for disregarding her avowed eclectic intentions in what follows. First, she does think that the choice is exhaustive in case (a), and this is all I deal with in this part. Second, my claim is that on an exhaustive choice between Theory and Simulation, Theory would be necessary. In this regard, Heal's position is merely a springboard. Third, Heal herself belies her avowed intentions with respect to (c) throughout her writing - in 'Replication and Functionalism': in her arguments from heuristic parsimony (in the *same* paper in which she avows the intention, 1995b, section 2 below); and in her argument from the Frame Problem (1996, p.28 below).

bulwark against the temptation of setting up Simulation as a straw-man. Third, it is Heal who poses the from-the-inside-idea challenge, so it is apt that a response should evolve out of her position. Fourth, as far as I am aware, she is the only writer who argues for Simulation from substantive content. The argument is a good motivation to develop an account of mental states *independent* of attribution considerations (which I do with respect to emotions in Chapter 2); and then to show that precisely considerations from substantive content count against Simulation (which I do in Chapter 3). Fifth, she offers the most lucid articulation of the often brandished claim that Theory is heuristically promiscuous. Finally, Heal only claims a limited scope for Simulation, which makes her position very strong: if Simulation has any chance of flourishing it must be on this narrowly and clearly circumscribed soil. Showing that Simulation is implausible even in such a best-case scenario, then, generates a strong argument in favour of Theory.

There are two stages to Heal's argument for Theory's heuristic inadequacy in contexts which involve prior knowledge of the subject's other current mental states:

1. showing that substantive content is indispensable to predictions;
2. arguing that Theory predictions cannot accommodate substantive content.

In section 1 I deal with her argument for the first claim, and with some considerations (of self-attribution) which favour her conclusion. In section 2 I impugn the second claim. In section 3 I show that Theory is necessary for predictions. Section 4 fleshes out the account of Theory sketched in Part I.

Section 1: Substantive content

Heal defines 'content' as "whatever is reported of a thought by the substantive descriptive words used in the that-clause specifying it" (1995b: 37). Her elucidation of the substantive nature of content involves two caveats. Firstly, we cannot ascribe to a subject a mental state with a certain content, without presupposing the subject's mastery of the concepts constituting the content and of their relations to other relevant concepts. Secondly, we must allow for plasticity²⁰ in a person's conceptual abilities. We should grant, in other words, the empiricist precept that the acquisition of new concepts is not reducible to a reconfiguration or "repackaging" (1995b: 37) of old concepts.

Heal's justification for the importance of substantive content to predictions turns on the consideration that "[t]hinking about thinking cannot be easier or simpler than the first level thinking itself. If it were, then, absurdly, I could short-cut my own laborious intellectual endeavours" (1995b: 36). Does this entail that *substantive* content, as characterised above, is

²⁰ I borrow this term from Peter Goldie (2000: 116).

necessary for predictions? A look at Tyler Burge's discussion of self-ascriptions²¹ may suggest that Heal's requirement of substantivity is too stringent.

In 'Our Entitlement to Self-Knowledge' Burge argues that self-ascriptions *inherit* the content of first-order thoughts. The notion of content that results is substantive enough to render self-attributions contentful and yet formal enough for their "self-verifying" (1996: 92) character not to be threatened by externalist considerations about concept-acquisition: whatever content was externalistically fixed at the first level would be carried over *individualistically* at the second. If we grant that predictions are a type of (forward-looking) attribution, then predictions could do with a similarly minimal notion of content. Once we have grasped the concepts involved in first-order thinking, the attribution/ prediction would make use of that grasp. This would mean that substantive content is not necessary for other-attributions.

It may be objected that the parallel will fail because Burge's concern is with *self*-ascriptions and what allows for the content to carry over to the second level is (as he himself urges) the fact that both the first- and second-order content are held from "a single point of view" (1996: 110). This, however, should have no effect on my point for two reasons. First, the "single point of view" is necessary to guarantee the *self-verifying* character of self ascriptions. Now, I am not arguing that other-ascriptions and predictions are privileged or infallible. Burge interests me here only for his offer of the possibility of first-order content being *inherited*. Second, it seems that once I have grasped the content in question, the attribution of that content to another will be from the same point of view, namely mine: *I* grasp that *p* and *I* attribute a similar grasp to *S*. The content gets carried over just as formally (though not infallibly, because *S*'s grasp of *p* may be different from mine) as in the self-ascription case.

The trouble is that even this formal notion of second-order content supports Heal's contention that substantive content is necessary for predictions: because the second-level content is inherited, it would (have to) be just as substantive as the first even though the *move* from the first to second level is formal. Furthermore, the second-level thought would fail without grasp of the concepts involved in the first. It looks like substantive content is necessary for attributions and predictions, after all. Returning to Heal's two caveats on content, this insight translates into the requirement that the ascriber needs to (a) have mastery of all concepts and conceptual relations requisite for the subject's thought; and (b) be equally prone to concept-plasticity in his attributions as the subject is in his first order-thoughts.

²¹ I am not suggesting that Heal accepts a Privileged Access view of self-knowledge. In fact she denies it (see section 3, 2. below). Rather, what I am suggesting is that even on a Privileged Access view substantive content is important in Heal's sense.

This complements my discussion of Gordon's position. I suggested there that Gordon would not get away with mere contagion, and must endow his simulator with all the concepts of the mental states causally operative in the simulation *mechanism*. Heal's idea translates into the further requirement that the simulator must possess all the concepts which figure in the content of the *input* state.

Section 2: Heuristic parsimony

Heal's next move is to argue that, considering her two caveats on substantive content, Theory must be heuristically promiscuous:

however elegantly the theory is axiomatised the fact remains that it is going to be enormously complex. Moreover we certainly cannot now formulate it explicitly. There should therefore be some reluctance to credit ourselves with knowing it (even implicitly) *unless there is no alternative account of how psychological explanation could work*. But there is an alternative." (1995a: 47, my italics).

And she proceeds to argue that Simulation is heuristically more economical than Theory in the context of predictions of mental states on the basis of the ascriber's knowledge of the subject's other mental states. She offers an example of trying to predict a fellow quantum physicist's (P's) solution to a certain problem, and urges that Theory needs to postulate an inordinately complex body of knowledge of others' mental states and their interrelations, in order to predict P's solution. The simulationist, by contrast, does not face the problematic lack of economy:

The basis for understanding a fellow quantum physicist ...is, she will say, the actual thinking about the subject matter which one person can do and which she can reasonably conjecture is the same as the one which the other has also done or will do (1995b: 40).

In this and the next section I argue for two claims:

- (a) The Simulationist approach seems to have the advantage of heuristic parsimony over Theory only because Heal either neglects certain of the elements involved in a prediction, or she smuggles in Theory's tools in the envisaged simulation (section 2).
- (b) Some Theory is necessary in a prediction, and once this is conceded it turns out that it is Simulation which is heuristically unwieldy (section 3).

There are two principal problems with the putative elegance of Simulation's negotiation of predicting P's solution.

1. *Assumptions about output and mechanism*

First, the view either assumes that there is a single solution to any given problem or (worse still) that both P and I would come up with the same solution in the face of a plethora of

possible solutions. I do not think this claim would be contended. If it is, I urge my opponent to consider Jon Elster's 'A Plea for Mechanisms' (1999). He cites there abundant evidence both from conflicting proverbs and conflicting statistics on outcomes of decision procedures. The data suggest not only that we can come up with different solutions to the same problems, but also that general assumptions about practical reasoning are insufficient to determine which one of two incompatible strategies/ decisions a subject would choose in any given context²².

If Heal did not assume that two people would come up with the same solution to a problem, it would not strike her as so obvious that I can simply simulate P and come up with *his* solution (and I need to come up with his solution, since what is at stake is a prediction of *his* state). This contributes another troublesome assumption to those in Part I: Simulation is not only making the unproblematic claim that we have shared *general* mechanisms for theoretical (and practical) reasoning. It also needs to assume that we share more fine-grained mechanisms, mechanisms for specific domains of reasoning and decision. But it is obvious that this will not do. Consider theoretical reasoning. The appropriate description of the mechanism should be something as general as Botterill's Perception and Inference principles (Part I, section 2 above). But by assuming that two agents would come up with the same solution to the same problem, the simulationist is, in effect, assuming that specific parts of this mechanism (like the one responsible for quantum physics) function identically. This cannot be right. The generality of the mechanisms implies that in any particular situation, even with the same input, as Elster points out, there is not necessarily a single rational thing to do or think.

2. Worries about input

The second, related, worry is that Heal seems to have forgotten that we are supposed to be predicting P's solution *on the basis of his other current mental states*. Should she have canvassed these two worries, Simulation would hardly be hailed as the paragon of economy. The availability of more than one solutions to a certain problem suggests that I have to take into account P's other mental states and broader psychological characteristics in a prediction of P: for instance, does he have a penchant for elegant solutions, is he likely to make some

²² It is important to note that this observation will not undermine my later claim that Theory is sufficient for emotion-attributions and -predictions, because my pet Theory encompasses *knowledge of character* as well as general assumptions (cf. Chapter 3, Part II).

unwarranted metaphysical assumptions for the sake of simplicity, and similar considerations²³. Notice that this is a problem even if we assume that our concepts are identical. The worries bring to light the fact that the same mental states with the same content can be related to, and hence generate, other mental states in vastly differing ways. This is something with which Heal agrees in 'Replication and Functionalism':

there is no hope of defining the idea of a particular psychological state...in isolation from other psychological notions. Such notions come as a package, *full understanding of any number of which requires a grip on the system as a whole* (1995a: 46, my italics).

For some reason, however, she ignores this complexity in her construal of Simulation. But if a heuristic model shuns *actual* complexity, of course it will be a simpler model! Whether it is going to be accurate, however, is another issue. So the promise of economy which Simulation seemed to proffer would be fulfilled at the dear price of ignoring the complexity and possibility for variety of mental-state interactions. Worse, such a promise would be self-defeating, since substantive content is at the heart of predictions

3. *Equal parsimony?*

Stich and Nichols²⁴ have argued that Simulation and Theory are equally (un)parsimonious, and so appeals to parsimony confer no advantage on either position (1995a: 137-8). The argument is that on both accounts prediction involves a two-level procedure - the first is the level of input, the second is the level of *processing* of input. So they conclude that Simulation gets its apparatus for the second level "for free" (1995a: 138), while Theory gets its apparatus for the first level for free. This argument seems to subvert my claim that Theory is more parsimonious than Simulation. If so, I must note that my argument from heuristic parsimony is mere methodological hygiene: it is important that I *respond* to Heal's challenge from heuristic parsimony in order to clear the way for the substantive arguments for Theory's necessity (Chapter 1) and sufficiency (Chapter 3). Stich's and Nichols's argument will do just as well for this purpose: as long as Heal cannot make her charge of Theory's heuristic promiscuity stick,

²³ Of course, as I have described these characteristics of P's, they do not amount to occurrent mental states. But this does not affect my argument: they constitute propensities for certain mental states and thus would (a) be causally responsible for, and thus relevant to, the predicted states; and hence (b) form part of the knowledge of mental states we (should) bring to the prediction. The objection from multiple solutions was inspired by Goldie's talk of substantive characterisation being necessary for predictions, as well as by his suggestion that more than one action/ emotional response would make sense in any given situation (2000:187).

²⁴ I am indebted to David Ryan (who arrived independently at Stich's and Nichols's insight) for persuading me that I should consider this option.

the way is clear for the substantive arguments. This is precisely what Stich and Nichols achieve. If neither Theory nor Simulation can claim heuristic-parsimony superiority, then heuristic parsimony does not strengthen or subvert arguments for either. Since neither my arguments for Theory's necessity nor for its sufficiency rely on considerations from heuristic parsimony, Stich's and Nichols's claim does not threaten, but favours my thesis.

Section 3: Theory's necessity for predictions

The consideration of the multiplicity of possible outcomes can be used not only to undermine Simulation's claim to economy but to foster the more substantive thesis that Theory is necessary for predictions. In this section I focus on two aspects of the input mental state: its substantive content, and its holistic character.

1. *Relevant similarity*

A step towards showing Theory's necessity for predictions is to point out that because of these complex possibilities, a second-level representation of the various contents is required together with a reliable and general account of the interrelations between types of content and of mental states.

The suggestion gains plausibility from Heal's use of "conjecture" in her description of Simulation (p. 21 above). What is this "reasonable" conjecture but a second-order judgement about my simulation (not) being a simulation of, or being relevantly similar to, the simulated thought-processes? By virtue of what is the relevant similarity gauged? As argued in Part I, the assumption that we share general decision-making mechanisms is not problematic. The problems are, rather, with the input state. There are two worries here neither of which can be redressed without recourse to metatheoretical considerations. Firstly, the ascriber must decide whether his mental states and the ascribed (input) mental states are *indeed* relevantly similar. That Heal herself sees this as necessary is evinced in 'Replication and Functionalism':

To get good results from the method I require only [!] that I have the ability to get myself into the same state as the person I wish to know about and that he and I *are in fact relevantly similar* (1995a: 47, my italics).

But (and this is the second issue and one which Heal ignores), surely the judgement must be in light of *criteria of relevance* which the simulator must *determine*. Certain mental states and characteristics would be more relevant than others to the prediction at hand, and it is

these that the ascriber needs to 'replicate' (in Heal's idiolect)²⁵. Both the decision about which features of the other are relevant to the simulation and the decision about whether the ascriber actually has replicated them, require metatheory. But once we have posited this metatheory do we need, in addition, to perform a simulation? Clearly not. We have the necessary links - when predicting a quantum physicist's solution, look for his character traits and his theoretical bias: is he going to fudge metaphysical issues, or is he a realist, is he interested in reconciling his theory with gravity, etc. Once we have answered these questions the specific answer to the problem that he will give becomes obvious (more of this in Chapter 3). Of course, one also needs to know some quantum physics. But this is consonant with Theory's model: we cannot make predictions in a domain whose concepts we do not have (see section 4 below).

As far as Heal is concerned, the relevant similarity criterion is very minimal and automatically satisfied:

Only one simple assumption is needed: that [people I simulate] are like me in being thinkers, that they possess the same fundamental cognitive capacities and propensities that I do (1995a: 47).

This of course is necessary for prediction, but is it sufficient to guarantee that a given simulation of a *particular mental state* is accurate? First, it should be noted that Heal's notion of similarity does not concern the *content* of the *input* state but the *mechanism* which we deploy in simulation. Recall that I distinguished the two notions of similarity in Part I (section 1) and argued that even if we grant that we share such mechanisms, we still need to establish that the *input* that we feed into the mechanism is relevantly similar to the *actual* mental state of the subject whom the input purports to replicate. I showed that the way to do this is by something like the analogical inference (since Gordon's account merely fudged the issue of input content). But the analogical inference, as even Goldman appreciates, is not a good recommendation for any model of theory that relies on it.

Heal circumvents the distinction between similarity of mechanism and of input throughout, but the following excerpt from 'Replication and Functionalism' suggests a way of making the mechanism criterion of relevant similarity sufficient for correct attributions:

on the view I maintain, one has no more access to the intrinsic nature of one's own thoughts than one does to the intrinsic nature of others'. Thinking about my own thoughts...is in my own case, as for others, to replicate - that is putting on a sort of performance, rather than being in possession of a certain kind of knowledge (1995a: 57).

Here is a suggestion then: it may turn out that if there is no asymmetry between the way we attribute mental states to ourselves and to others, all the relevant similarity needed is indeed as

²⁵ Heal uses 'replicate' synonymously with 'simulate'. I use it more loosely, to denote 'reproduce'.

minimal as Heal suggests. This idea is seductive for two reasons. First, it does seem that if (a) our self-attributions do not advert to metatheoretical considerations about content; and (b) our self-attribution mechanism is the same as our other-attribution mechanism, then metatheory becomes redundant for other-attributions. Second, this vision illuminates Heal's charge of promiscuity: Theory's requirement that we develop - over and above concept-manipulation - an ability to represent such manipulation becomes implausible, since it does not make sense in the first-person case.

Should we accept this etiolated criterion of relevant similarity? Well, it looks like (a) and (b) are a reiteration of the problem rather than its solution.

a. Recall that champions of Simulation argue from the putatively obvious fact that we engage perpetually in simulation when thinking and deciding about the future. Heal is no exception. She cites hypothetical reasoning (1998: 91) as the paradigm of simulation *and* self-attribution. And the move from here to claim (a) is clear. As Stich and Nichols have pointed out (1995b: 98-9), however, the supposition that our thought about the non-actual works by simulation simply begs the question at issue. This seems right. Heal makes no attempt to argue for the claim, she just assumes it.

b. Even if we grant that we simulate the non-actual, however, a fresh problem arises: are thinking about *my* non-actual mental states and about *another's* mental states sufficiently similar? This brings me to claim (b) above. Heal offers no argument for the alleged symmetry between self- and other-attributions. The following considerations should weaken her fondness of (b).

First, it is unclear that any account which relies on a symmetry between self- and other-attributions has much appeal. We have the prephilosophical intuition that there is something privileged about our self-attributions: I do not need to observe my own circumstances and my behaviour to know that I want a glass of water; whereas I do need all this to know that another does. Of course, many philosophers (most notably, G. Ryle) have argued that this is precisely how we self-attribute. The debate is by no means settled, however, and this consideration serves at least to show that Heal owes us an argument here. Now, this may not seem a problem for Heal, since she is concerned with attributions on the basis not of behaviour but of knowledge of the other's antecedent mental states. But actually it is a problem, because we also have the intuition of asymmetry when it comes to self-attributions on the basis of other mental states. Recall the discussion of Tyler Burge in section 1. When I attribute to myself the thought that *p*, the attribution inherits the first-order content *p* automatically and self-verifyingly, because the content is held 'from a single point of view'.

When I attribute to S the thought that p, in one sense the content is still held from a single point of view (mine) but in this case the second-order thought is not self-verifying. Because the attribution is of the form 'I think that *he* thinks that p', there is no guarantee that 'I' and 'he' will have the same concepts figuring in the content p, and so no guarantee that 'I' am right about his believing that p. The only thing that would guarantee this is if I *knew* that my relevant concepts are identical to his or if they are not, I must adjust for relevant differences. But this is a piece of Theory, which needs to be *used* in a simulation, otherwise the prediction is arbitrary (as I argued against Gordon).

Second, this seemingly gratuitous objection may grow quite substantive if we consider that for (a) and (b) to confer advantage on Simulation, the self-attributing and other-attributing mechanisms must be equally *reliable*. Seeing that they are not, it becomes unclear why Simulation (supposedly the same method for both) works so well in the self-case and not in the other-case. Heal may object that in the other-case we never know enough of the other's mental states to run a perfect simulation. This route, however, is not open to her, because all she thinks that Simulation requires, by way of relevant similarity, is that I know that the other is a thinker.

2. *Demonstrative specification of simulated mental state*

Now, Heal acknowledges the need for

allow[ing] somewhere for the idea of different personalities, for different styles of thinking and for non-rational influences on thinking.

But, she continues,

It is not clear what shape such additions to the core replication process would take. But there is no reason to suppose that they would take the form of the proposed functionalist-style theory (1995a: 48-9).

Heal envisages Simulation's opponent as a functionalist. Of course, by the way I have fleshed out theory, it should be clear that I am not committed to functionalism. So the question is whether "such additions" can come from anywhere but Theory. I claim not. But Heal would have to deny the privilege not only to functionalism but to any version of Theory. The reason for this is her demonstrative account of mental-state individuation.

In her account of what is involved in understanding judgements about others' mental states, Heal concedes the need for a less minimal criterion of similarity, namely that the simulation not be mere contagion, but a pretence:

If I pretend something I must know that this is what I do and I must, in some sense, know what I pretend (1995b: 44-45).

This is consonant with what I have said so far. But, she continues,

I may have no explicit way of specifying what I pretend other than by pointing to the pretence itself; *it is the demonstrated individual features of that actual performance which will then carry the weight of the specification* (1995b: 45, my italics).

As noted earlier, Gordon adopts a similar view of ascriptions: "To ascribe to O a belief that p is to assert that p within the context of a simulation of O" (1995a: 60). So this is a general account of ascription and a way of avoiding objections from relevant similarity. But the immediate puzzle these accounts generate is that if the specification of a mental state is in terms of the pretence itself, then there will obviously be no way of discerning whether the pretence is indeed a pretence, that is, (a) a conscious and intentional act (which is necessary for something to count as a pretence²⁶) rather than contagion (which I argued is insufficient for predictions, Part I, section 1); and (b) whether my pretended state indeed replicates the state I am ascribing to a subject. Such judgements of similarity can only be made on the basis of specification of the relevant mental state, made *independently* of the pretence, and hence need to be handled by a metatheory.

I have argued in this section that even if Simulation is the appropriate account of predictions, some Theory would need to be invoked at least at the beginning of the envisaged simulation procedure. The argument appealed to two aspects of the input states: substantive content, and the holistic character of mental states. The first requires Theory in order to provide *criteria* of relevant similarity between the content of my input states and the content of the simulated states; and to establish whether my input states are indeed relevantly similar to the simulated states. For both these tasks we need reliable principles which connect content to concepts and to other relevant mental states. The holistic character of the mental makes Theory necessary for predictions because it implies that certain mental states would be more relevant than others for the prediction at hand. Theory is necessary to establish what these states are and how they should be used in a simulation.

It may be objected that my criticism of Simulation is based on Heal's very minimal and incomplete characterisation of what is involved in Simulation. My criticism, however, does not depend on Heal's construal. Rather, it depends on the consideration that any more complete account would have to incorporate the subject's other mental states, and hence would need - for

²⁶ I am aware that G. Currie has disputed that either consciousness or intention is necessary for pretence (1995: 163). About consciousness he may be right. But if he impugns the intentionality of pretence, then he seems to have changed the subject. In any case *Heal* cannot agree with him (and this is all that counts here), because a non-intended pretence would collapse, for her, into contagion.

the sake of completeness - to advert to a metatheory which provides criteria for (a) which concept/ mental state belongs to which *type* of content/ mental state, and (b) what counts as relevant similarities between the ascriber's and the ascribed mental states.

Section 4: More relevance

The most forceful argument that Heal (1996) musters against Theory is based precisely on considerations from relevance. She argues that Theory faces a version of the Frame Problem. Recall that I argued that Theory is needed for determining which mental states are *relevant* to a simulation. Heal turns the tables and notices that, considering the enormous body of psychological knowledge that Theory postulates, it is a puzzle how an agent determines which precepts are relevant to a particular prediction. She claims that the Theorist has only one choice: to postulate a meta-body of information, "a general theory of relevance" (1996: 81), by which posit he either invites an infinite regress or has solved the Frame Problem. The Simulationist, by contrast, has no such worries: when we use our own theoretical and practical apparatuses, they naturally weed out the non-salient aspects of a situation, just as they do when they work in the first-person. And of course, the big advantage is that they do not need to sift through reams of theory and discover which of its elements are relevant to a prediction.

Heal's challenge should be taken seriously, not least because it threatens to show that Simulation is necessary for predictions. And it should be clear by now that even though I claim that I am only showing that Theory is necessary in this chapter, which is compatible with Simulation's also being necessary, it would be a sorry triumph for Theory if its most important work is done by Simulation. So let us take Heal seriously. Heal's claim has two parts:

1. Simulation, unalloyed, offers a better account of how we deal with relevance;
2. Even if Theory is the apposite approach to predictions, Simulation will need to be invoked when it comes to determining which part of the body of psychological knowledge is relevant to a particular prediction.

1. *Simulation's superiority*

First, Heal's argument for the superiority of Simulation will only have unequivocal force for predictions of mental states on the basis of knowledge of the other's *situation* and *behaviour* (though I will dispute even that in Chapter 3). The simulator can look at the situation and let his own cognitive apparatus determine what is salient about it, and then proceed with the simulation on the basis of that. No representation of the situation is needed, so it looks like Simulation is vindicated. If we are concerned with predictions on the basis of *other mental states*, however, (as we *are* here) it is unclear that Simulation would be of much help in

determining relevance. As I have argued above, the simulator cannot just automatically feed the other's mental states in his own mechanism. He would need a *metarepresentation* of these mental states, in order to determine whether they are relevantly similar to his own. As I have also argued, this amounts to Theory's necessity for predictions²⁷. So Simulation cannot be *sufficient* (at least) for predictions of mental states on the basis of knowledge of other mental states.

2. *Simulation's necessity*

Next comes the claim that Simulation would be *necessary* for a Theory-based prediction. It is founded, unsurprisingly on the massiveness of the envisaged body of information. I agree that the infinite regress is a genuine problem, but do not agree that the Theorist must invoke a metatheory which deals with relevance.

a. First, the massiveness of the body of knowledge is dubious. If we accept Botterill's suggestion, there are some very basic general principles at the core of theory. So at this level, the relevance-procedure is very simple: if action is at stake, use Action Principle; if predicting beliefs on the basis of knowledge of the subject's situation, use Perception Principle; if predicting an inference, use Inference Principle. But, of course, it cannot be this simple, because substantive content as well as mental holism will be involved, so we need a whole lot of knowledge about the subject himself, as well as rules of the kind 'Egoists tend to act in x sort of way in y situations'; 'Evangelists tend to ignore certain evidence when making inferences of type z', and so on. It is these *empirical* generalisations that are most susceptible to Heal's charge.

b. Let us agree with Heal that in many cases we determine relevance by reference to ourselves. Does this mean that we need to be involved in a simulation? I contend, no. Part of the folk's theory is knowledge of what 'I' would do in a particular situation, as well as knowledge (or self-deception) about how I compare to normal rational beings, to egoists, and to Evangelists. This knowledge may play a central role in determining which elements are relevant to any particular prediction. But so what? I am not feeding my mechanisms with pretend input, I am not identifying with the other. Rather, I notice aspects of a situation, have a *representation* of how 'I' am, a *representation* of how the other is, and *infer* that if his eyes are working

²⁷ I am aware that Sterelny (2001: 230) has claimed that (what he calls) "metarepresentation" is a much more minimal notion than Theory. It would follow that showing that someone engages in representation does not entail that he is engaging in Theory. I think, however, that he is using the term 'metarepresentation' in a much more minimal sense (considering that animals, according to him, engage in it). On my use of the term, representation involves *conceptualisation* and so of necessity Theory.

properly, and he likes flowers, then he will be enraptured with this rose; because a rose is a flower and this one is a particularly gorgeous specimen. Of course, *I* have to *see* that this rose is gorgeous, but hopefully Theory-friends are not expected to walk the earth blindfolded just to vindicate Simulation.

The envisaged Theory-method is the same for more complex predictions: if I am predicting a quantum physicist's solution to a problem, then I would have to solve the problem myself. But my solution is just that, mine, and not a *prediction* of his solution. I am just solving a problem. My solution can only become a prediction of his solution if I then compare our personalities, our conceptual sophistication, and our methodologies. My solving the problem only seems like a simulation of another because of the complexity of the domain: there is just no way of predicting a quantum solution except by doing the problem myself. But in our ordinary predictions of each other we do not need to go through the feelings and thoughts of the other in order to predict what he will feel, think, or do.

Conclusion

In this part of the chapter, I have argued that Heal's considerations of substantive content count against Simulation, and in fact favour Theory's necessity for predictions of mental states on the basis of the ascriber's knowledge of the subject's other mental states. Heal has, however, contributed an important addition to the sketch of Theory, offered at the end of Part I: the central role that the ascriber's representation of himself may play in determining which piece of theory he must use in a particular prediction. Theory, then, is still modeled on Botterill's analogy with a Lakatosian core, where

- (a) the central tenets are the Action Principle, the Perception Principle, and the Inference Principle, which capture the conceptual links amongst mental states on the one hand, and amongst mental states, situational input, and action on the other;
- (b) the periphery consists of general rules of thumb based on empirical knowledge about how things generally tend to work, which is partly based on how things generally tend to work *with me*;
- (c) and the relevance of a particular set of precepts from this database is determined by personal salience, as well as a representation of how the ascriber compares with others.

This then is the version of Theory necessary for predictions. In the next part, I corroborate this necessity by considering Iago's predictions of Othello's emotions. The discussion will be useful both for supplementing the vague examples argued from so far; and for introducing the account of emotions which I develop in Chapter 2.

Part III: Theory and the Green Ey'd Monster

Introduction

Friends of Simulation are fond of invoking literature as the final (auspicious) arbiter between their position and Theory. In this section, I accept the challenge, and show that although literature entices us into empathy through an appeal to our imagination, the appeal would be fruitless if imaginative situations did not rely for their effect on a Theory we are assumed to possess²⁸. For this purpose I examine closely a specific case to test Simulation. I focus on Iago's strategy of heart-breaking Othello into jealousy²⁹, and argue that what Iago is doing there is fraught with theory, from Iago's, Othello's, and the audience's points of view. My argument unfolds as follows. In section 1 I examine some of Iago's soliloquies in order to furnish *prima facie* plausibility for the claim that he is using Theory in predicting that Othello will be jealous under some (contrived by Iago) circumstances. In section 2 I argue that Iago is incapable of the kind of jealousy he provokes in Othello, or of Othello's antecedent love. In section 3 I show that Iago lacks empathetic imagination³⁰. Now, assuming that we grant that Iago's predictions of the effects of his strategies on Othello are correct (and if we doubt that, we doubt that Othello ever got jealous), and that Iago had no Simulation tools (capacity for being in relevantly similar states not/ through empathy), then it is obvious that he could not have used Simulation for the predictions. Finally, I argue throughout the section that the reason the audience painfully anticipates Iago's success is not that we simulate Othello's psychology when Iago sketches his stratagems, but because we recognise at each step the accumulative effect they will have. The recognition, I contend, is based on our (Theory-driven) knowledge about what would elicit jealousy in a person like Othello.

²⁸ One way of showing this would be to consider the use literature makes of metaphor, and argue that metaphors are the author's short and sure way of automatically conjuring relevant chapters of the Theory he supposes we possess. This would show that our imaginative activities themselves depend on Theory. Such analysis, however, may prove to be rather vague and leave us as unsatisfied as insufficiently characterised examples (like Heal's one about the quantum physicist) do.

²⁹ Jon Elster (1999: 109-111) also considers Iago in his analysis of emotions, but uses him for very different purposes: to establish what sort of emotions he is undergoing, rather than to analyse his apparatus for predictions of Othello's emotions.

³⁰ I owe the distinction between imagination *simpliciter* and empathetic imagination to Sam Naidu and Dan Wylie. They both suggested that despite standard arguments for the concomitance of evil and lack of imagination *simpliciter*, anyone involved in such immaculate deception as Iago is must have some imagination. I agree with both this claim and their aversion to the association of evil with lack of imagination. Milton's Satan in *Paradise Lost* (despite Milton's intentions) is more imaginative than the Trinity and the angels put together.

Section 1: Iago's apparent use of Theory

This is how Iago describes his strategy:

1. [Othello] holds me well;
The better shall my purpose work on him.
How, how?...
2. After some time to abuse Othello's ear
That [Cassio] is too familiar with [Othello's] wife.
3. [Cassio] hath a person and a smooth dispose
To be suspected - framed to make women false.
4. The Moor is of a free and open nature
That thinks men honest that but seem to be so;
And will as tenderly be led by the nose
As asses are. (1.3.384-396)

The soliloquy, as syllogistic as it is, forcefully suggests that Iago is wielding a Theory about the necessary conditions for duping someone into jealousy. A look at (what I have numbered as) the individual conditions will illuminate this claim.

(1) seems a piece of theory about the deceiver's *trustworthiness* being felicitous for the deception. Indeed this sentiment is echoed by Othello himself at several points in the play. For instance, when Iago begins to sow doubt in Othello's mind, by feigning bewilderment and echoing everything Othello says, Othello falls for the tactic, because (he says):

such things in a false disloyal knave
Are tricks of custom; but in a man that's just
They are close delations, working from the heart
That passion cannot rule (3.3.125-128).

Notice that not only does Iago's reasoning seem like a piece of theory but Othello's adumbration of it is law-like and generalised.

(2) spells out the strategy to make someone jealous: to induce the *belief necessary* for jealousy. Now, in the next chapter I will argue that belief is not necessary for emotions, and some weaker doxastic notion will suffice. But for the moment it is enough to note the syllogistic character of Othello's reasoning about the necessity of belief:

I'll see before I doubt; when I doubt, prove;
And on the proof, there is no more but this -
Away at once with love or jealousy! (3.3.194-196)

(3) seems a piece of theory about (a) what makes a person more likely to "make women false" - being attractive (and not a hundred lines earlier, Iago adds youth); and (b) that such attractiveness in the rival lends credibility to the adultery.

(4) seems a piece of theory about trustfulness and generosity in the one deceived being necessary for the deception. In the next scene Iago elaborates on this criterion. He plans, he says, to:

Make the Moor thank me, love me, and reward me,
For making him egregiously an ass,
And practising upon his peace and quiet
Even to madness (2.2.301-304).

I have treated each condition cautiously, saying that it *seems* to evince Iago's use of a theory. This provides only initial plausibility to my thesis. It may be that Iago is simulating Othello's thoughts, and alerting the audience to the individual stages of his simulation. The question which arises, however, is why the audience would need such sign-posting. If simulation is indeed the way we attribute mental states, why is it that Shakespeare - the master of dramatic impact - did not write Iago's soliloquies as a simulation of Othello's thoughts? Why is it that what Iago is running through in every soliloquy are general *rules* of deception, fleshed out by ethological considerations about the protagonists at hand? A step towards answering these questions is to argue that Iago is using a theory. Otherwise it becomes a puzzle why the audience needs such sign-posting if simulation is the *actual*, and most *economical* and *efficacious* way we attribute mental states.

Another consideration is that Iago is incapable of simulation because he does not meet a necessary condition for simulation. If my arguments from relevant similarity in Part II succeeded, the capacity for being in relevantly similar states as the person simulated is such a necessary condition. He fails the criterion for two reasons: (a) he is incapable of love and hence the sort of jealousy into which he whips up Othello (section 2); and (b) his imagination is extremely limited, too limited in fact to run a simulation of those of Othello's mental states causally responsible for his jealousy (section 3).

Section 2: Iago's incapacity for love and (Othello's kind of) jealousy

Which psychological features of Othello's cause his jealousy? Let us grant that Iago's strategy is successful and take our cue from him. It looks like Othello's love for Desdemona, together with his trustfulness (which is responsible for his belief that she is unfaithful) do the trick. That Iago is incapable of trustfulness is obvious. I turn then to his incapacity for love.

1. Love

Shortly before Othello embarks on his final self-eulogy, he urges the nobles of Venice and Cyprus to believe that “naught I did in hate, but all in honour” (5.2.298), and describes himself as “one who lov’d not wisely but too well” (5.2.347). Both these statements are consonant with the image the audience has formed of him. By contrast, everything Iago does in the play is, *by his own admission*, done “in hate”. The play opens with his hatred of Othello (1.1.6); reveals his various personal hatreds; and shows his last words to his wife, before he stabs her to death, to be “Filth, thou liest” (5.2.234).

His descriptions of love do not fare much better in disclosing a capacity for love. “It is merely a lust of the blood and a permission of the will”, he says (1.3.333). It may be objected that he expresses this sentiment to Roderigo whom he is also trying to dupe, and thus this is an unreliable symptom of his attitude to love. There is, however, enough evidence in the text that this *is* how he feels about love. For example, the only love he ever expresses (in a soliloquy) is, interestingly, for Desdemona (2.1.285). But it is not, as it soon transpires, what we would normally understand by the notion. He tells us that it is not “absolute lust” (which is the closest he comes to our notion of love), “though paraventure/ I stand accountant of as great a sin”. More important, what *causes* his love is not even solely “absolute lust” but a desire

to diet my revenge,
For that I do suspect the lustful Moor
Hath leap’d into my seat (2.1.286-289).

2. Jealousy

This passage also makes it clear that Iago’s notion of jealousy is very different from Othello’s. The first time he mentions his suspicion that Emilia may have been unfaithful with Othello, he comments:

I know not if’t be true;
Yet I for mere suspicion in that kind,
Will do as if for surety (1.3.382-384).

The striking thing about this ‘jealousy’ is that it utterly lacks either Othello’s passion (for that matter *any* passion) or Othello’s (or anyone’s) concern with the main object of jealousy. In fact, Emilia’s name is not mentioned at all. We *infer* that Iago is jealous of *her* because of his use of “office” when he voices (in his characteristically poetic manner) his suspicion that Othello “twixt my sheets/ Has done my office” (1.3.381-2). This lack of concern with what is the appropriate object of jealousy, together with a sense that Iago uses his ‘jealousy’ as a mere excuse for injuring others is adumbrated in a passage in which he claims that jealousy “Doth like a poisonous mineral gnaw my inwards” and immediately proceeds to say:

nothing can nor shall content my soul
Till I'm even'd with him wife for wife;
Or failing so, yet that I put the Moor
At least into a jealousy so strong
that judgement cannot cure (3.1.292-296, my italics)

Again, the putative object of jealousy neither is the cause of jealousy nor figures in his torment as anything more than the appropriate coin in a barter-deal. By contrast, Othello's immediate concern is with Desdemona, and only secondarily he acquires a hatred for Cassio as the rival.

These considerations suggest that Iago is incapable of *experiencing* Othello's thoughts and emotions, and hence cast serious doubt on whether he is capable of *simulating* them. To use Simulation speak, it looks like Iago does not have the mechanisms relevant for loving and jealous thoughts.

3. Competence with concepts of love and jealousy

Before I confirm this suspicion, it should be noted that despite his incapacity for love and for the relevant jealousy, Iago understands both *concepts*³¹ perfectly well and knows how to handle them in diagnoses (which, of course, is why he succeeds with Othello).

a. He ascribes love with perfect ease to Cassio (2.1.283), Desdemona, and Othello ("the Moor is of a constant, loving, noble nature", 2.3.283), and understands that "his soul is... enfetter'd to her love" (2.3.334). Moreover, he can discriminate between different kinds of love and knows that the kind of love Desdemona bears Cassio is "apt and of great credit" (2.1.281).

b. His competence extends to the concept of jealousy - he ascribes it very accurately and knows just how it works:

Trifles light as air
Are to the jealous confirmations strong
As proofs of holy writ (3.3.326-328).

What we have so far, then, is that a person who lacks the capacity for certain thoughts and emotions can, nonetheless, ascribe them in a substantive and nuanced way to others. Moreover, his reasoning about these thoughts and emotions sounds very much like he is engaging in Theory. This provides great plausibility for the thesis that we indeed use theory rather than Simulation in predictions. The Simulation champion, however, is still not vanquished. He can

³¹ I am not suggesting that Iago has *full* mastery of the relevant concepts, but that he has sufficient mastery in order to attribute competently. I am indebted to Eusebius McKaiser for pointing out that claiming that Iago has full master would be dodgy.

argue that the apparent Theory-use is for dramatic purposes, while Iago's incapacity for love and the right jealousy does not preclude him from simulating these emotions. Concerning the former, as I have already mentioned, it would not make sense for a dramatist to use Theory-driven soliloquies if the way our attributions work is through simulation. What of Iago's incapacity for these emotions not precluding him from simulating them? The only plausible Simulationist answer would be that Iago transcends his own character *through imagination*, and thus simulates Othello.

Section 3: Iago's 'imagination'

Imagination is of great concern to the play. When brought before the nobles of Venice to account for the "witchcraft" (as Brabantio calls it, 1.3.64) which won him Desdemona's love, Othello tells movingly of how he wooed her with story-telling (1.3.130-169). Throughout the play he is depicted as a man of vast imaginative power. Iago, by contrast, emerges as a person of rather limited imagination. It is hard to pinpoint exactly how he creates this impression, beyond the fact that he thinks about others in a rather formulaic manner. People for him never transcend certain character-types: Roderigo is stupid, thus exploitable; Cassio is noble, exposes Iago's own baseness and thus killable; Othello is trusting, thus exploitable; women are Venetian women, thus mendacious and generally ruled by their appetites (3.3.232-237)³². This line of thought, unfortunately, seems rather circular: Iago is using a Theory, because he cannot simulate; and he cannot simulate for his lack of imagination which we in turn diagnose on the basis of his formulaic (Theory-driven) treatment of others' complexity.

But how else do we gauge someone's imaginative capacities if not by reference to his world-view and to his engagement with others' world-views? I have already said something about the former (Iago's notions of love) while the latter proved rather circular. Let me return to his world-view. Iago's is a simple universe:

If the balance of our lives had not one scale of reason to poise another of sensuality, *the blood and baseness of our natures* would conduct us to most preposterous conclusions. But we have reason to cool our *raging motions, our carnal stings, our unbitted lusts* (1.3.318-331, my italics).

The italicised parts here suggest Iago's exhaustive view of emotions as a base and embarrassing fact about us. There is nowhere here, or elsewhere, a tribute to the complexity or

³² This specific statement of the promiscuous and deceptive ways of Venetian women is addressed to Othello, and is part of the overall stratagem to elicit his belief that Desdemona is unfaithful. This fact may make it doubtful that this is indeed Iago's perception of women. His treatment of Bianca and his wife, however, allays such doubts. For example, when Emilia brings him the coveted handkerchief with the words "I have a thing for you", he answers "It is a common thing" (3.3.305-307). The pun typifies his attitude and behaviour towards women throughout the play.

beauty of our emotional lives. Still, this may not be a failure of his imagination in the needed sense: it does not preclude a capacity for empathy, though it tells strongly against its likelihood.

Perhaps one way of arguing for Iago's imputed lack of empathetic imagination³³ is by pointing out that he treats others solely as means to his own ends. This line of reasoning, however, confuses empathy with sympathy. The fact that Iago uses everyone in the play may show that he lacks sympathy, but sympathy is obviously not necessary for empathy³⁴, and hence not necessary for simulation.

There are two points in the play which crystallise Iago's incapacity for empathetic imagination: one decisive and one inconclusive (but as Iago knows inconclusive evidence "may help to thicken other proofs/ that do demonstrate thinly", 3.3.435-436). The inconclusive episode is when Roderigo (who genuinely loves Desdemona) threatens to drown for love, and Iago exclaims "Well, if thou dost, I shall never love thee after it. *Why*, thou silly gentleman!" Roderigo's answer is perfectly intelligible to anyone who has loved: "It is silliness to live when to live is torment" (1.3.305-310). Of course, Iago is playing a game here, but this moment is confirmed later as a moment of possible incomprehension on Iago's part later: when Othello threatens to kill Desdemona, Iago suggests that he "strangle her in her bed, even the bed she hath contaminated" (4.1.204). Now this will emerge as proof that Iago is incapable of empathetic imagination in light of the following two considerations:

- (a) Iago does not want Desdemona killed;
- (b) his suggestion has such impact that this is indeed how Othello kills her;

These two considerations suggest that Iago fails to appreciate the power of Othello's love and jealousy. The reason, I propose is that he cannot conceive of what is actually involved in (including the fatal consequences of) the *passionate experience* of such emotions. This is what shows that Iago's incapacity for emotions of such scope preclude him from their imaginative simulation.

The only objection I can envisage for this argument is that a) may be a dubious claim. I am working on the assumption that Iago adheres to his intentions revealed in his several soliloquies. It may be urged, however, that there is no proof that Iago did not change his plan and decide to harm Desdemona while he was at it. Two considerations make this reading less plausible than mine:

- (a) There is no reason why this detail should change in Iago's strategy, and especially not without Shakespeare's alerting the audience to the change (considering Iago shares every single step of his envisaged strategy);

³³ I owe this idea to Sam Naidu.

³⁴ Goldie has also suggested this rather obvious point in 1999 and 2000.

(b) the episode in which Iago suggests this develops at high emotional intensity on Othello's part. It is also clear from its increasing momentum that Iago is improvising in inflaming Othello's passion (in unison with his intention to drive him to madness, 2.2.301-304). Both these points suggest that Iago is kindling Othello's imagination for his own purposes without appreciating the possible fatal consequences. And I think that anyone who does not want such consequences *and* fails to appreciate their inevitability lacks a notion of *experiencing* love and jealousy, and hence the requisite empathy/ mechanism for simulating such passions in the substantive-content sort of way required by Simulationists like Heal.

Conclusion

The present chapter developed three classes of arguments. Part I offered general doubts about Simulation's envisaged predictions. It was exposed as being forced to exploit the pernicious analogical inference, or to ignore substantive content, and/ or to beg the question against Theory. Part II grappled with Heal's version of Simulation. It was argued that Theory is not only *more parsimonious* than Simulation, but that it is *necessary* for predictions, since predictions need to make use of substantive content. Finally, Part III offered an actual example of predictions. Theory's necessity was evinced in Iago's apparent use of a Theory in predicting Othello's emotions, and his lack of the relevant mechanisms which should have been operant if Simulation was correct. In light of these three conclusions, Theory emerges as necessary for predictions of any mental state on the basis of knowledge of the subject's other mental states. The next chapter develops an analysis of the nature of emotions, by way of preparation for exploiting the conclusions of Chapter I in the context of emotion-attributions and predictions (which I do in Chapter 3).

CHAPTER 2

The secret of our emotions never lies in the bare object, but in its subtle relations to our own past: no wonder the secret escapes the unsympathizing observer, who might as well put on his spectacles to discern odours (George Eliot)³⁵.

Introduction

In this chapter I argue for a *sui generis* model of emotions according to which (i) emotions' objects and causes are concern-based construals; and (ii) emotions *qua* attitudes are (a) complex states embedded in a narrative structure, (b) characterised in terms of their object, their expressive behaviour, and their phenomenology. I proceed as follows. In Part I, I impugn reductive conceptions of emotion. In Part II, I introduce Ronald de Sousa's account of emotions' objects. In sections 3 to 5 I pursue the object of emotions and elaborate on its individual aspects by developing Roberts's conception of concern-based construals. I argue that such construals are the *typical causes* and *necessary objects* of emotions. Part III focuses on emotions as *attitudes*. I discuss emotions' expressive behaviour and their phenomenology, and argue that they partly constitute emotions (sections 1 and 2). In section 3 I introduce the notion of narrative structures by way of evolving criteria for judging emotions in/appropriate. Finally, I develop a Relational Schema for narratives (section 4). Except for section 3, where I offer a more systematic analysis of rationality, I address throughout the chapter, issues of rationality alongside constitutive issues.

Part I: Why we cannot reduce emotions to other mental states

Introduction

One of the central questions in the debate concerning emotions is whether emotions are *sui generis*, or are reducible to other mental states³⁶. William James (1884), for example, has argued that emotions are nothing but our awareness of our bodily changes. Others, inspired by him (and the problems his account faces) have construed emotions more broadly, as general feelings or sensations. A third group of philosophers (most prominently Robert Solomon, 1980)

³⁵ *Adam Bede*, bk. II, XVIII, p. 196.

³⁶ I ignore the separate question about whether emotions form a proper class of mental states. Many writers (e.g. Griffiths, 1997; Rorty, 1980) have argued that they do not. All that is relevant for my purposes is that emotions are a more or less unified *explanatory* category. That they are such a category is evinced by our everyday practice. Once this is conceded, my account needs to show that emotions are mental states which have some central characteristics in common. This will be achieved in Parts II and III.

have argued that emotions are nothing but evaluative judgements. A more nuanced view (Joel Marks's, 1982) is that emotions reduce to a belief/ judgement-desire complex. By contrast, several philosophers have argued that emotions are *sui generis*. Amongst these are Amélie Rorty (1980), John Deigh (1994), Michael Stocker (1983, 1996), Ronald de Sousa (1987), Robert C. Roberts (1988), and Peter Goldie (2000, 2002).

The standard arguments *sui-generis* exponents offer against the physiological and sensations accounts is that proprioceptions and sensations are not intentional states in the required sense - they are directed, if at all, at the body. The paradigmatic objects of emotions, by contrast, are in the world. So, while it is generally agreed that emotions are embodied, and thus involve both proprioceptive and feelings awareness, this awareness does not exhaust their complexity.

Concerning the other reductivists (generally known as cognitivists), there are three main classes of arguments against them³⁷:

1. arguments juxtaposing the involuntary nature of belief and the voluntary component in emotions;
2. arguments based on belief's insufficiency for emotions;
3. arguments based on the insight that belief is not necessary for emotion.

Section 1: Doxastic involuntarism

Arguments of the first class conclude that emotion cannot be reduced to belief(-desire complexes) on the grounds that beliefs are involuntary while emotions can be controlled (in some sense)³⁸. Now, this argument lacks force. I am a doxastic involuntarist and think we have some control over emotions, but am not persuaded. The reason is that if we look at *occurrences* of beliefs and emotions the lack of control is equal - I cannot just believe/ stop believing "at will" (Williams, 1973) or "just like that" (Bennett, 1990) and similarly I cannot become/ stop being angry at will or just like that. But in both the doxastic and emotion cases, I can cultivate certain attitudes which can lead to my adoption of a particular belief³⁹ or my propensity for particular emotions. So this line of argument need not discomfit the reductivist. But I think we can do much better with the next two classes of argument.

³⁷ These three types are summarised by Roberts (1988). I develop them considerably, since as they stand they are rather schematic and confer at best *prima facie* plausibility on the *sui generis* view.

³⁸ Roberts (1988), Goldie (2000: 72). Though de Sousa does not develop such an argument to my knowledge, he agrees with both the partial voluntariness of emotion (e.g. 1987: 11) and is a doxastic involuntarist (282).

³⁹ Thus James Montmarquet (1993) argues for our duty to cultivate doxastic virtues.

Section 2: Belief is not sufficient for emotion

Arguments, against cognitivism, of the second class are based on the consideration that belief is not sufficient for emotion. I can judge something dangerous without being afraid. There are two possible responses for the cognitivist: to add something to belief in order to make it sufficient; or to deny that belief is insufficient.

1. *Adding on desire*

Along the first lines, Joel Marks argues that what is missing in the above example is my desire not to be harmed. So an emotion is a belief or judgement plus a relevant desire towards the content of the doxastic state. The belief-desire complex view will not fare much better, however. As Roberts points out (1988), if beliefs and desires can be dispositional, then as long as I am not attending to either my belief that something is dangerous or my desire not to be harmed, or both, I can still believe something dangerous without fearing it. And if it is objected that what is involved is a judgement (rather than a dispositional belief) we can still argue that there are cases in which the desire is dispositional and so I can still judge something dangerous without fearing it. The notion of construal in Part II will clarify this point.

2. *Denying that belief is insufficient*

The alternative response to the contention that belief is insufficient for emotion, best developed by Solomon, is to deny the claim. Solomon argues that emotions are evaluative judgements⁴⁰ and when responding to the putative insufficiency of judgments, he writes:

But an emotion is never a single judgment but a system of judgements, and although one might well make one of several judgments of the system without having the emotion, my claim is that one cannot make *all* of them and not have the emotion (1980: 275).

A few lines later, as if himself unpersuaded by this *ad hoc* move, he adverts to Marks's line of reasoning and decides to add a desire to the judgement. Then, finally, he resorts to a notion of caring about the content of the judgement:

One might make a judgment...in an impersonal and uninvolved way, without caring one way or the other. But an emotional (set of) judgment(s) is necessarily personal and involved (1980: 276).

The reduction still does not go through. Only the first two of Solomon's three moves are reductive, and neither works. The first does not work, because if judgement is insufficient for

⁴⁰ Actually, he later modifies this claim to "emotions are defined primarily by their constitutive judgements" (1980: 274), but the response I sketch here is to the claim that judgements are insufficient for emotion.

emotion, multiplying the judgements is not going to help (unless we invoke some weird form of emergence). The second move does not work for the reasons Marks's view does not. Finally, the only promising move is the third, but it amounts either to circularly importing emotions at the level of involvement/ care, or to denying that emotions are simply a belief/ judgement-desire complex. When I develop the notion of concern-based construals, it will become evident that this latter move is perfectly compatible with a *sui generis* view of emotions.

Section 3: Belief is not necessary for emotion

The final, and heftiest, argument against reductions of emotions points out that belief is not necessary for emotion. It seems that one can experience fear of something without believing that the object of fear is dangerous. Stocker (1983) offers the example of fear of flying despite knowledge that statistically it is far less dangerous than driving. Similar examples abound, and all that the cognitivist can do is to postulate *ad hoc* unconscious beliefs which make little sense of oft-encountered statements like 'I *know* I am being silly to fear it, but I can't help it'. Of course, the cognitivist can respond to this by denying that such a situation is possible. The fact that I fear flying and act fearfully on planes, the envisaged response goes⁴¹, indicates that I actually believe flying to be dangerous, despite my genuine (but alas deluded) assertions that I believe flying harmless.

The problem with this response is the parenthetical comment. We have to assume in every instance of an emotion with an avowed conflicting belief that we have a case of either self-deception or utter irrationality - an agent embracing without qualms Moore's paradox. Self-deception on such a large scale does not seem to be an option. Moreover, the case does not look like self-deception: the subject of such emotions knows that the avowed belief does not square with whatever doxastic state is supposed to underpin the avowed emotion. If a belief is necessary for emotion, then, we are left with agents regularly indulging in Moore's paradox. Such a vision is undesirable for three reasons.

1. Charitable considerations

First, it breaches the Principle of Charity. This is fine on occasion and with some individuals. The phenomenon of (what Rorty calls, 1980: 104) "conservation of emotions", however, is by no means so rare. On the contrary, it seems that emotions get their bad press precisely from their resistance to reasons and to other relevant *beliefs*! Such resistance is not nearly so pervasive amongst our beliefs (and certainly not in the Moore's paradox form, which is why it is called a paradox). The first reason for not accepting the envisaged objection, then, is that it

⁴¹ I am indebted to Francis Williamson for persuading me that I should take this objection seriously.

suggests that agents engage promiscuously in Moore's paradox, and are thus far less rational than our epistemology presupposes. There would also be an irony here: because a state (fear of flying) is not responsive to reasons and other beliefs (flying is harmless), we call it a belief (flying is dangerous)!

2. *The constitutive role of actions and inferences*

Second, if it is objected that such appeals to charity are dubious, we may notice the assumption that my opponent is making. He argues that since the person who claims that he believes flying harmless, *behaves* (where this includes the emotion and ensuing actions) as if he believes it dangerous, we should ascribe to him the belief that it is dangerous. My opponent's assumption then is that action is partly constitutive of a belief. Now, I agree with this wholeheartedly. Unfortunately, the insight subverts my opponent's position. The constitutive relationship between action and belief is a *normative* one: that is, a "rational believer" cannot be said to believe that *p* unless he acted as if *p* is true at least most of the time (Stoneham, 2000: 16).

But this is not the whole story. Another normative commitment of a rational believer's believing that *p*, is that he make at least many of the right *inferences* from *p*⁴². My claim, then, is that if my opponent appeals to actions as the auspicious arbiter of whether someone has a belief, he is judging the presence of a belief on the basis of a believer's normative commitments. But as soon as he does that, he must take into consideration inferences too (since actions and inferences are *jointly* sufficient for a belief). But now consider the person who fears flying. He mostly *acts* as if flying is dangerous, but most of his *inferences* are governed by the proposition that flying is harmless (which is why so many people who fear flying plan holidays, buy airplane tickets, get on a plane, and argue with their emotion throughout the journey). My claim then is that as soon as my opponent appeals to action as the arbiter of belief, he has no reason to conclude that either the belief that flying is harmless nor his putative belief that it is not, counts as a belief at all.

3. *Considerations from epistemology*

This brings me to the final reason why we should not accept my opponent's contention that the person who fears flying actually *believes* flying dangerous. Several writers, most prominently Cheshire Calhoun, have noted that arguments like my opponent's ultimately frustrate our epistemological projects by entailing a binary model of belief⁴³. Calhoun argues that rather than

⁴² Tom Stoneham argues for these two normative commitments persuasively in his 'Self-Knowledge' (2000: 16).

⁴³ I am indebted for this phrase to Tom Martin.

forcing emotions into dogmatic norms of doxastic commitment, we should consider the very real phenomenon of conserved emotions as a source of illumination for epistemology. So, she offers the example of a person who believes not to be a homophobe, being awkward even revolted when it transpires that a friend is gay. Calhoun uses the example to generate a distinction in the way we can hold a belief: “theoretically” as opposed to “evidentially”⁴⁴. It is not my aim to develop the distinction here (I think it is too limiting in any case⁴⁵). I am rather interested in the line of reasoning, which is sound: very often we appreciate that there is a conflict between a genuinely held belief and the doxastic commitment underpinning an emotion. Any resolution of the conflict by either denying that the avowed belief is a belief or reducing the complex emotion to a belief, must be too facile and unsatisfactory. Rather, because of the pervasiveness of such states, we should focus on developing an account of emotion, which acknowledges their reality and complexity, and would thus allow for doxastic nuance that will ultimately solve some central epistemic problems.

Conclusion

We have, then, at least three good reasons for rejecting my opponent’s response to the claim that belief is not necessary for an emotion: the Principle of Charity; considerations from the normative role actions *and* inferences play in constituting belief; and finally that the opponent’s position presupposes (and ultimately supports) a binary model of belief. Belief, we are entitled to conclude, is not necessary for emotion. I also argued that it is not sufficient. To these arguments we might add de Sousa’s insight (1987: 165) that emotions have very different formal objects from both belief and desire⁴⁶. Belief and desire have unique and single formal objects (truth, and the good/ desirable respectively). Each emotion type, by contrast, has a different formal object (e.g. fear’s object is the dangerous, love’s - the lovable, anger’s - the culpably offensive, and so on). If these arguments worked, we should no longer be tempted to reduce emotions to beliefs or to belief-desire complexes.

⁴⁴ R. S. Dillon, who argues for a parallel distinction between “intellectual understanding” and “experiential understanding” (1997: 239), has taken the project further by focusing on the notion of self-respect in the context of women who believe they have every reason for self-esteem and yet cannot bring their emotions to square with such judgements. She has urged that if we discredit their avowed beliefs, we may miss the importance of the conflict: it mirrors an important tension in our society between the official policy of equality and pervasive non-egalitarian practices.

⁴⁵ Rorty argues persuasively for *five* varieties of doxastic commitment (1980: 112-3).

⁴⁶ I am adding this as an afterthought not because it is unimportant, but rather because the way the point is normally developed is somewhat question-begging: if I think that emotions reduce to beliefs, then I will obviously think that they share belief’s formal object, despite the apparent variety of emotions’ formal objects. Noticing the variety of emotions’ formal objects, then, only has force once we have persuaded the reductivist that belief is neither necessary nor sufficient for emotion.

I have not, I appreciate, totally eliminated the possibility that emotions are beliefs-desires-*feelings* complexes. Of course, the claim that belief is not necessary for emotion makes this alternative, at least *prima facie*, unlikely. I will have more to say about it in Part III (section 3). For the moment I will just sketch Goldie's response, which will have weight only once I have fleshed out his proposal and supplemented it by a more detailed account of the objects of emotion (cf. Part III, section 2). Goldie argues against (what he calls) this "add-on" (2000: 40) view on the grounds that it is unfaithful to our experience of emotions.

Phenomenologically, emotions are not just dispassionate judgements with a feeling tucked on. Rather an emotion is a unified complex experience irreducible to (the sum of) its individual components. His notion of "feeling towards" which is "thinking of with feeling" (2000:19) is intended to provide the conceptual underpinnings for this phenomenal intuition. It captures the special way emotions are intentional - they are not merely directed at the same *objects* at which beliefs are directed, and they are not the same kind of *attitude* as beliefs or desires.

It is time, then to develop a positive *sui generis* account of emotions. I start with de Sousa's account of emotions' objects. I then supplement it with Roberts's notion of concern-based construals. Finally, I tackle the issue of what sort of attitude emotions are, through a journey into Goldie's conceptions of "feeling towards" and of narrative structure. In the process, I will be concerned with developing criteria for the rationality or appropriateness of emotions. The reason I do this *in the course* of developing an account of emotions, rather than before or after the account is in place, is that it is both hard and undesirable to disentangle the question of what states emotions are, from the question of how we judge an emotion appropriate. It is undesirable, because inevitably we need to import some normative criteria at the descriptive level in order to make sure that certain phenomena classify as emotions. So, for example, I said earlier that the *normative* commitments of a *rational* believer - appropriate actions and inferences - partly *constitute* our notion of belief. Similarly, we need to develop an account of emotions based on their well-behaved exemplars, and then see how their inappropriate family-members relate to the decorous ones⁴⁷. This does involve, however, a

⁴⁷ This intuition has its theoretical underpinnings in what de Sousa calls the principle of "the priority of success". To elucidate it, he cites Millikan:

If...mental intentional states...are members of proper function [sic] or 'biological' categories, then they are...intentional states not by virtue of their powers but by virtue of what they are supposed to do yet perhaps cannot do...[I]f...we push the analogy with biological categories, only true beliefs are capable of performing the defining functions of beliefs. We will then be free to look for the defining attributes of beliefs among relations between true beliefs and the actual world outside. False beliefs will then appear merely as things that were 'supposed to' have had such and such relations to the outside world (cited in de Sousa, 1987: 114).

range of assumptions about emotional seemliness, and so a conception of their rationality must be developed alongside a constitutive account.

Part II: Emotions and their objects

Introduction

Ronald de Sousa, in his *The Rationality of Emotion*, proposes that emotions are individuated in terms of the *relation* between the emotion *type*, the *subject* and the *object* of the emotion in question. In this and the next two sections I focus on the object of emotions in an attempt to flesh out his proposal. In section 1 I introduce his account. In section 2 I focus on his notion of motivating aspect, which grounds de Sousa's normative account of emotions by virtue of distinguishing between emotions' causes and their objects. In sections 3 I develop Roberts's account of emotions as serious concern-based construals. I argue that emotions' causes should be viewed in light of construals. In section 4 I argue that concern-based construals are the *typical causes* and the *necessary objects* of emotions. When understood in the context of the correct Relational Schema, they also turn out to be sufficient.

Section 1: The Relational Schema

De Sousa distinguishes between six aspects of the object of emotion, which are normally conflated: target, focal properties, motivating aspect, cause, aim, and propositional object. The *target* is an "actual particular" at which the emotion is directed (1987: 115). *Focal properties* are those of the target's properties which are a conscious focus of the subject's attention and his grounds for the emotion he experiences (1987: 116). The *motivating aspect* is in standard cases (a) an *actual* property of the target, which (b) is a *focal* property, and (c) relates *causally* to the emotion. The difference between *cause* and *motivating aspect* is that the former figures in causal explanations of the emotion, while the latter provides reasons explanations. The *aim* of an emotion is defined in terms of its biological goal which is, in turn, manifested in the emotion's "immediate expressive behaviour" (1987: 120). The intuition is that we do things in the world out of emotion and the sorts of things we do are constrained by the biological origins and functions of the specific emotion. The *propositional object* specifies in propositional form the subject's ground for the emotion.

This then is the "relational schema" which individuates an emotion:

(RS) *Relational Schema*.

R (*S t f a c m p*)

where R stands for an emotion type, *S* is the subject, *t* the target, *f* the focal property, *a* the motivating aspect (which in the standard case is identical with *f*), *c* the cause, *m* the aim, and *p* the proposition specifying the ground of [the emotion in question] (1987: 126).

These parameters are not all necessary for all emotions. For example, forward-looking emotions like hope and worry will typically have only a propositional object (137-9). Love, on the other hand, may not have a propositional object at all (I 'love *someone*'. I do not 'love *that...*'). The point of the schema is that (apart from the aim which does not fit neatly and to which I will return in Part III) it canvasses all the possible aspects of an emotional object. It thus provides criteria for individuation. Emotion *types* are fixed by the characteristic "polyadicity" (the typical number of parameters) of the relevant schema. *Particular* emotions are individuated in terms of the specific values of the parameters.

The schema also provides criteria of intelligibility and appropriateness. An emotion is *intelligible* if (1987:122):

(a) its causal explanation coincides with its reason explanation, that is, the cause coincides with the target's focal properties;

(b) *as far as the agent is concerned*, the focal property is an instantiation of the formal object of the emotion; for example, if I am frightened by a snake (target) I need to see, say, its poisonous fangs (focal property) as an instance of the dangerous (formal object of fear).

An emotion is *appropriate* if it is intelligible and "if the target *actually* has a focal property in virtue of which the formal object fits the target" (1987:122, my italics). So intelligibility is a function of the values of the parameters of the Relational Schema, while appropriateness is a function of these parameters and their relation to the world.

The next three sections grapple with three of the parameters of the Relational Schema: motivating aspect (section 2); cause (section 3); object as captured by the notion of focal properties (section 4). In Part III, I return to the aim and argue for including phenomenology in the Relational Schema.

Section 2: Motivating aspects

First, I turn to a closer examination of de Sousa's notion of a motivating aspect. I have already mentioned that it needs to be a focal property which (1) is *causally*⁴⁸ involved in the emotion;

⁴⁸ "First Causal Condition. A causal connection between the focal property and the occurrence of the emotion is a necessary condition for the former to be a motivating aspect" (1987:117).

(2) *rationalises*⁴⁹ the emotion, and (3) is an *actual*⁵⁰ property of the target. These constitute the set of necessary and jointly sufficient conditions for a property to count as a motivating aspect. I tackle each in turn.

1. *The First Causal Condition*

That the motivating aspect causes the emotion is necessary so as to avoid cases in which the putative focal property has nothing to do with the aetiology of the emotion. For example, I become increasingly angry at someone for wearing purple nail polish, but actually my anger is caused by a combination of too much coffee and his owing me money. The nail polish then plays no causal role in my anger, and we would obviously not want to say that it motivates the anger.

2. *The Intelligibility Condition*

That the emotion is rationalised by its motivating aspect ensures that something like having had too much coffee cannot be a *reason* for my anger. So de Sousa offers an example of Wendy's despising Bernie ostensibly for his musical tastes but actually, unbeknownst to her, his voice reminds her of her hated grandmother's, and this is the true cause of her contempt. It cannot, however, constitute a reason, since it is - motivationally - just as arbitrary as having overdosed on coffee (and Wendy would admit this were she to realise the true cause of her emotion). It is not clear from *The Rationality of Emotion* what the scope of this condition is. De Sousa claims that it is a constraint on what can count as a *motivating aspect*. When he revisits the example, however, he is reluctant to call Wendy's attitude an *emotion*, because "the discovery of [its] real cause ... must almost certainly render the word 'contempt' inapplicable" (1987: 119). The observation points to the already mentioned difficulty of developing an account of emotion independent of normative considerations. I will return to this issue when I tackle extremely irrational emotions (Part III, section 4).

3. *The Second Causal Condition*

The requirement that motivating aspects be actual properties of the target needs a little more attention than the other two conditions. De Sousa motivates it by an example of emotion directed at what he calls an "illusory focus" (1987: 119). Bernie admires Wendy for having been a virtuoso violinist as a child. Wendy, however, has never been a virtuoso violinist. The

⁴⁹ "*Intelligibility Condition*. Motivating aspects must be *rationally related* to the emotion they cause, in the sense that they must constitute *intelligible rationalizations* for the emotion" (1987:118, italics in original).

⁵⁰ "*Second Causal Condition*. For a focal property to be a motivating aspect, it must be an actual property of the target" (1987: 120).

question is what has caused Bernie's admiration. It cannot be Wendy's virtuoso (focal) properties because she lacks them. On the other hand, it cannot be Bernie's belief that she has this property, since if that is the case, either the causes of our emotions never coincide with their objects, or all our emotions are caused by our beliefs about the world rather than by the world itself. This seems to force us in the absurd situation of, for example, fearing our beliefs about dangerous objects rather than the objects themselves. And so, de Sousa writes, if emotions are caused by our beliefs,

then you would always be wrong about the cause of your emotion regardless of your beliefs about it. For the content of your belief is that the relevant focal property is the cause of your emotion. But on the present proposal that cannot be true, for the cause of an emotion is always the belief itself. Nor can that belief figure as its own content, because that would make it viciously self-referential. It would become the belief that the emotion was caused by the belief that the emotion was caused by the belief...(1987: 120).

Now, there are two ways of interpreting this passage: as constituting either a descriptive or a normative claim. The descriptive interpretation would amount to the denial that beliefs can cause emotions. The normative interpretation would amount to the claim that *false* beliefs cannot constitute motivating aspects and hence cannot cause *appropriate* emotions. There are three reasons for interpreting the passage as a normative claim, all based on considerations from charity. The first is that no one would seriously maintain that beliefs cannot cause emotions on the grounds that if they did, then we are always afraid, or jealous, or angry with our relevant belief rather than with the state of affairs it captures. To impute this view to de Sousa would be uncharitable. Secondly, it would be unfair to ascribe the view to him, because it would commit him to a conception of experience on which we have unmediated access to the world: if beliefs cannot cause emotions, then the causes of emotions are always entities in the world. This is plainly inimical to de Sousa's acknowledgement that a subject has a point of view in the world⁵¹. The acknowledgement is evinced in his distinction between the target of the emotion (an object/ situation in the world) and its focal properties (which are irreducibly subjective). Finally, the reason we should read the passage as placing normative constraints on emotions is that it is aimed at elucidating the notion of motivating aspect which itself is partly normative. If the passage was supposed to elucidate focal properties - a descriptive notion - then we would be entitled to read it descriptively.

⁵¹ In case a tighter definition of this rather obvious notion of a point of view is needed here, Peter Goldie's is a start: he defines it as "the point of view of a conscious person, capable of thoughts and feelings, and able to engage in theoretical and practical reasoning" (2000: 1). And, though the definition does not make it explicit, this involves a *subjective perspective* on the world, understood both spatially and psychologically.

The suggestion that the normative interpretation is the apposite one gains weight from de Sousa's comparison of the Second Causal Condition to one of the necessary conditions for inferential knowledge. There, two criteria need to be satisfied:

- (a) that the belief be not merely justified but *caused* by the antecedent premises/ beliefs (this corresponds to the First Causal Condition);
- (b) that "the belief in the premises must have been caused by the fact" (1987: 120).

And, de Sousa urges,

[s]imilarly with emotions. The grounds of the emotion correspond to the premises of an inferential belief: they must cause the emotion, or else they are non-starters (ibid.).

The Second Causal Condition then "allows the world to be relevant to emotion. It provides the element of genuine relation guessed at by our preanalytic intuition" (ibid.).

Taking the inferential knowledge analogy seriously, however, would seem to make uncertain the actual (as opposed to envisaged) scope of the Second Causal Condition. The analogy would seem to betoken descriptive ambitions: when the premises of a final belief are not causally grounded in the world, the belief does not amount to knowledge; similarly, the analogy implies, a putative emotion not ultimately grounded in the world would be no emotion at all. Again, I suggest charity here. The analogy strictly applies to *appropriate* emotions rather than to emotions *tout court*. This makes sense considering that knowledge is a partially normative notion: it is a kind of appropriate (and true) belief⁵².

So far I have discussed de Sousa's notion of motivating aspect. A motivating aspect is an actual, focal property, causally involved in the emotion. (When I return to appropriateness in Part III, I will dispute the actuality condition.) As I have already noted the notion of a motivating aspect is a normative one. It is time to turn to its descriptive constituents - cause and focal properties. I start with causes.

Section 3: Causes as serious concern-based construals

De Sousa is quite articulate about what *cannot* be the cause of an *appropriate* emotion, but not very helpful with what *are* the causes of emotions in general. One possible answer here is that our emotions are caused by objects and situations in the world. But if this means that the most *proximate* cause of emotion is in the world, then we are presupposing unmediated access to the

⁵² I am aware that Gettier (1962) has challenged this conception of knowledge as true justified belief. For present purposes, however, the challenge does not matter, since this is the model on which de Sousa is working and that is all I am interested in here.



world, which is clearly false. We must, then, make sense of this mediatedness about our experience. I have already argued in Part I that claiming that beliefs are the necessary causes of emotions is equally undesirable: we seem to emote about a whole range of things which we do not (quite) *believe*. We need, then, some notion of doxastic commitment which is flexible enough to cover a whole range of doxastic states from full-blooded belief on the one extreme, to anaemic doxastic states expressed in statements like ‘This situation just *seems* dangerous to me, though I do not believe it is’ (where the ‘*seems*’ is sufficient to motivate). In this section I develop Roberts’s notion of construals and argue that it is best suited to cover this entire range of doxastic nuances as well as to provide the most accurate and useful conception of emotions’ causes.

1. *Serious concern-based construals*

According to Roberts emotions are “serious concern-based construals”. Roberts defines a construal as “a mental event or state in which one thing is grasped as something else” (1988: 190). The paradigm for construals is the perceptual construal of the duck-rabbit as a duck. But this is just an analogy. Construing is an “in terms of” relation [which] can have as its terms any of the following: A perception, a thought, an image, a concept” (ibid). Most of our mental activity consists of such construals. So, for example, seeing someone as (reminiscent of) my long-dead friend involves a construal in terms of an image (he looks like my friend), or in terms of a concept (he is just as witty), or in terms of a thought (he makes me smell the flowers in the same way). And, of course, the notion of a construal implies the notion of salience: it involves “dwelling on or attending to, or at a minimum holding onto, some aspect [of the object of the construal]” (1988: 187). So, my seeing someone as my long-dead friend involves *focusing* on some aspects of this someone and holding onto those aspects. The “holding onto” here can cover the full doxastic gamut: from ‘I *believe* that he is my dead friend’ to ‘He just seems so like him that I will play with him as if he were, even though I know he isn’t’.

Now, according to Roberts, what distinguishes emotions from other mental states is that they are “concern-based” construals. By concern Roberts means “desires and aversions, and the attachments and interests from which many of our desires and aversions derive” (1988: 202). So to be angry with someone “is to construe him as having culpably offended in some matter of concern to me” (1988: 205). Notice that a belief is not necessary. What plays the role of a belief is that the construal is a “serious” one, that is “for the construer, it has the appearance of truth whether or not she would affirm the truth of the construal” (1988: 191).

I will return to this notion of seriousness in the next sub-section. For the moment, it should be noted that Roberts is not arguing for an atomistic view of the emotions in the way a

cognitivist argues that an emotion just is a belief and a suitable desire. An emotion is not a concern *plus* a construal, rather it is a concern-imbued construal:

a construal is not an interpretation laid over a neutrally perceived object,
but a *characterisation of the object, a way the object presents itself*'
(1988:192, italics in original, my emphasis).

The underlined phrase brings us head on to the problem of the emotional object. What does it mean for an object to present itself? In the duck-rabbit picture it seems, from a reason-explanatory point of view, pretty arbitrary whether the duck or rabbit presents itself (though of course the presentation has a causal history - if I am an ornithologist, perhaps I am more likely to see the duck). Are emotional construals equally arbitrary? It would seem that if certain aspects of an object or situation just present themselves to us, then yes. But of course that is the point of the "concern-based" part of the concern-based construal. The concern in question *rationalises*, as well as makes it more likely, that we attend to certain aspects of the object rather than others. And what sort of concerns we have will be shaped by biological and social factors as well as by our beliefs, desires, character traits, and of course emotions. I will further elucidate both the notions of concern and of the special way concerns imbue our construals presently (sub-section 3.).

I used the word "shaped" in the second to last sentence, because Roberts uses it throughout. In fact he never mentions causation. So, it is time to return to the question of what causes our emotions. On the present proposal it is the object of the concern-based construal. But this is not very helpful since it amounts to the claim that the cause of the emotion is the object of the emotion. So, say that X is an object in the world, at which my emotion is directed. And say that X_c is the object shaped by construal, and X_{cc} is the object shaped by a concern-based construal. Which of these three is the cause of my emotion? It cannot be X itself since I have no access to X except through a certain filter of construals (otherwise we have unmediated access to the world). It cannot be X_c because a dispassionate construal is insufficient for an emotion (this is what motivated talk of concern in the first place). X_{cc} is the only option left. But there seems to be a problem here. If an emotion is a concern-based construal and X_{cc} is the *product* of such a construal (that is, X would not appear in the X_{cc} way to me unless the concern-based construal was already made), then the emotion seems to have been caused by an object which would not exist without the emotion in the first place! We must, then, either not identify the emotion with the concern-based construal or not identify the cause of the emotion with X_{cc}.

My suggestion is that we opt for the former. *Pace* Roberts, an emotion is not itself a concern-based construal, but is a mental state caused by the object 'produced' by such a

construal⁵³. So certain concerns make me focus on particular aspects of a situation or object or event (to construe it, in other words, in a certain way) which then cause, and are the object of, my emotion. To put it in de Sousa's idiolect, certain concerns make me focus on certain aspects of the target (focal properties), which then cause the emotion.

The proposal is a refinement of de Sousa's model in the sense that (a) it spells out what the causes of emotions are; and (b) does so in a way that makes appropriate emotions formally continuous with inappropriate ones. The causes of both are of the same type - a serious concern-based construal. The difference between emotions with a real focus and emotions with illusory focus is one of degree: in the first case, the construal corresponds closely to the real properties of the target, while in the second, the construal has either exaggerated dramatically, or created, the properties of the target. On de Sousa's model illusory-focus emotions were a puzzle because he never spells out what the causes of our emotions are; and those with illusory focus were suggested to be different *in kind* from appropriately caused emotions. The problem was, in other words, that de Sousa implied that appropriate emotions were caused by the *actual* properties of the target, rather than by an *accurate construal* of them. On the present proposal the actual properties of a target never cause anything since they need to go through a construal filter (no matter how minimal). And the construal filter is something that both rational and irrational emotions must go through. This is not some barbarous call to subjectivism. It is just an accurate picture of our experience - only the things that I notice about an object can play a causal role in my mental life. In the perception case, for example, the construal filter is my eyesight, if it is poor then for me certain visual properties of an object at a certain distance just do not exist, and hence cannot cause anything *in me*.

The proposal, then, is that emotions are caused by serious concern-based construals. I am now in a better position to return to two promised elucidations: of *seriousness* and of construals' being *imbued* with concern.

2. *Seriousness*

Recall that Roberts defined a serious construal as one which has, "for the construer...the appearance of truth whether or not she would affirm the truth of the construal" (1988: 191). A

⁵³ The suggestion has the additional advantage of avoiding having to grapple with a standard objection against Roberts. The objection is that he defines an emotion as a concern-based construal, but if concerns themselves are emotions, then the definition is trivial. Now, because I do not define an emotion *as* a concern-based construal, but have it caused by such construals, I can afford that some concerns can turn out to be emotions: no one denies that emotions can be caused by other emotions. Of course, it had better not be the case that *all* concerns are emotions, but we can avoid this by noting that concerns can be *unemotional* biological, cultural, aesthetic, or ethical desires and aversions. I am indebted to David Ryan for contributing ethical desires, and to Thad Metz for worrying that Roberts's definition may be circular.

good start on elucidating seriousness is an example that Roberts offers, of being angry with his two-year-old for spilling ketchup over her new dress:

even though I hold a theory of moral development that rules out her being culpable for this heinous act, I construe the situation as one in which a responsible agent has culpably offended. She *looks* guilty to me (1988: 201, emphasis in original).

The all-too-familiar example suggests that we do *in some sense* which is “compelling” (to use Roberts’s word, 1988:201) take, at least for the moment, the Xcc type object to be the real X. Depending on how closely Xcc corresponds to the real object, the emotion-causing can be rational or utterly irrational. And the jocular tone of the passage suggests that, surely, in this case, Dad is being unreasonable even by his own lights (even if these lights switch on only half an hour after the episode).

Perhaps such seriousness comes in degrees. The suggestion is strengthened by the consideration that it is in virtue of the seriousness involved that the construal acquires both doxastic status and causal powers (consider: someone else’s two-year-old can look - as opposed to *look* - guilty, and Roberts will presumably not feel angry). And it would be neat to establish a correspondence between the degree of seriousness and the degree of doxastic commitment, since construals were introduced in order to cover a range of doxastic commitments. I doubt that it is fruitful to cast the difference of degree in terms of a difference in *intensity* of seriousness. Such an approach would seem to reintroduce the original problem of accounting for degrees of doxastic commitment.

Seriousness, I contend, can only be a useful notion here if we make it a function of motivation, and, more importantly, if we time-index it: the stronger its motivational force and the longer the seriousness persists, the heavier the doxastic commitment involved in the construal. So, Roberts is only serious for a little while in his construal of his two-year-old as guilty: she *looks* guilty only while the ketchup stain is around him; for pretty much the rest of the time she *looks* lovable, cherubic, musical and generally love- and pride-inducing (if reports on paternal love are correct). And I suspect that how long his seriousness lasts with respect to construing her as guilty would depend on the temporal as well as ethological significance of his relevant concerns: the concern to have his daughter in church, or wherever they are going, in her cleanest and newest dress is presumably not as central as to see her grow into a wonderful person⁵⁴. So I said that someone else’s child would not upset Roberts, because he would not

⁵⁴ On the other hand, if Roberts could not afford to buy her other dresses and/or one of his central concerns in life was with ‘what people would think’, the seriousness may persist for much longer. Such seems to be the case with parents who are perpetually angry with, or generally antagonistic to, their child for their own failures.

take the construal seriously. But it would seem that the reason he would not is that the relevant concerns are not in place. This is beginning to sound rather circular, though: we are looking for something in serious concern-based construals to account for their doxastic proteanism; we latch onto the *seriousness* component as the crucial criterion of flexibility; and then explain the seriousness itself in terms of *concerns*. On the other hand, perhaps such concept-incest is symptomatic of the account's genuinely *sui generis* nature. So let me pursue the concern track a bit further.

3. *Concern-imbuement and magnetizing dispositions*

The idea of construals' being concern-imbued is best elucidated by Rorty's notion of "magnetizing dispositions". In her 'Explaining Emotions' Rorty distinguishes between two types of emotion-causes: "immediate" and "significant". What I have been discussing so far have been immediate causes. The significant cause, by contrast, is "the set of events - the entire causal history - that explains the efficacy of the immediate or precipitating cause" (1980: 106). Rorty offers an example of a man (Jonah) resenting his female boss even though all his colleagues rightly respect and like her greatly. Rorty traces the significant cause of the resentment (rather boringly) to Jonah's fear of his domineering mother, whom he perceives as having repeatedly tried to strangle him as a child by tying (too tightly) warm (but too itchy) scarves around his neck?! Further, the phobic construal itself derives from Jonah's paternal grandfather, who clearly did not like his daughter-in-law, and with whom Jonah stayed at the time when his brother was born (and Jonah was feeling vulnerable and anxious about his mother's love and attention in any case). The *immediate* cause of Jonah's resentment of his boss, then, is his construal (though Rorty does not use the term) of her as tyrannical and threatening. The *significant* cause is a set of concerns (again not Rorty's term) with the threat of powerful women in general; with self-preservation both literal and in terms of self-esteem; and so on. The important point is that the significant cause is not the set of events identified in an objective manner (or "extensionally", 1980: 109), but rather as identified by their significance for the individual.

The set of events in their significance for the subject shape what Rorty calls the subject's "magnetizing dispositions". These are dispositions "to gravitate towards and to create conditions which spring other dispositions" (1980: 106). More particularly and importantly, such dispositions determine our "habits of selective *attention* and *interpretation*" (1980: 106, 108, my italics). And so, Rorty writes:

It is because significant causes often produce magnetizing dispositions that they are successful in explaining the efficacy of the immediate causes of the emotion: they explain not only the response but the

tendencies to *structure experience* in ways that will elicit that characteristic response (1980: 107, my italics).

This notion of structuring experience in such a way that it engenders certain emotional responses can be used to flesh out Roberts's notion of a construal's being *imbued* with concern. Concerns shape our dispositions in the way that magnetizing dispositions are said to do. And the reason the notion of imbue, rather than causality, is more apt here, is that what we are telling is not a straight causal story. Suppose, to take a prosaic example, I have had food poisoning for the last three days. My persistent nausea colours the way I apprehend all food - even food that I perceived as attractive a few days ago now *looks* disgusting. Now, of course, in one sense this is a simple causal story: nausea causes disgust. But the reason imbue is more appropriate here is that it is not the case that I see the food in a neutral sort of way first (as strawberries, say) and then judge it to be attractive four days ago and repulsive now. Rather, four days ago the strawberries presented themselves to me as strawberries-to-be-relished and now appear to me as strawberries-to-be-vomited. And it is in this way that the concerns which form our magnetizing dispositions and derive from particular sets of events in our lives, imbue our construals which then cause emotions.

The phenomenon of concern-imbued construals has long been familiar to novelists. Dickens depicts it movingly in *Martin Chuzzlewit*. When Tom Pinch discovers that Mr. Pecksniff has never been the paragon of moral rectitude and generosity that Tom for years thought him to be, the simplest of objects around him - before always infused with Pecksniff's image - acquire a new and unfamiliar emptiness. When Tom goes to his bedroom to pack his things, Dickens relates Tom's associations with the room and comments:

At any other time he would have parted from it with a pang, thinking of all he had learnt there, of the many hours he had passed there; for the love of his very dreams. But there was no Pecksniff; there never had been a Pecksniff, and *the unreality of Pecksniff extended itself to the chamber...* (*Martin Chuzzlewit*, 31.431, my italics).

The omnipotence of concern-imbue culminates in Salisbury - always before a source of wonder and pleasure to Tom - in which he arrives destitute and Pecksniffless. Dickens comments significantly:

Oh! What a different town Salisbury was in Tom Pinch's eyes to be sure, when the substantive Pecksniff of his heart melted away into an idle dream! He possessed the *same* faith in the wonderful shops, the *same* intensified appreciation of the mystery and wickedness of the place...and yet it was not the old city nor anything like it (36. 475, my italics).

Again, Dickens describes the market through which Tom is walking as the same market (he uses the word “same” over ten times in the paragraph) and yet, he says, “it was strangely changed to Tom”.

The passages simultaneously capture imbuelement neatly and testify to its causal powers: Tom is precluded from feeling certain emotions towards his very memories, by the sheer force of his construal of the world around him as devoid of Pecksniff and hence of meaning. The imbuelement, poignantly, extends to the most trivial of Tom’s daily activities:

Tom had so long been used to steep the Pecksniff of his fancy in his tea, and spread him on his toast, and take him as a relish with his beer, that he made but a poor breakfast on the first morning after his expulsion (36. 375-6).

It should by now be clear why I needed to invoke concern for the notion of degrees of seriousness to get off the ground. In typical cases, it seems that the *nature* and *origin* of the concern will determine how serious the seriousness involved is, and how long it can last. Roberts’s seriousness lasts only a while because stained dresses figure very little in his self- and daughter-concerns and conception. Tom’s concern with moral rectitude, by contrast, is an essential part of his self- and other-conception. The seriousness such concerns are prone to producing is unlikely to vanish in a matter of years, let alone minutes. This is because concerns *qua* magnetizing dispositions derive their causal/ imbuelement clout from our individual histories: Tom’s ideal of Pecksniff is the very locus of Tom’s self-conception, which is why, once he is disabused, the entire world around him is permeated with a sense of loss, is itself incomplete (an incompleteness rendered powerfully by Dickens’s elegiac tone in the cited passages as well as throughout the chapters from which they come). When I introduce narratives, I will return to this conception of individual history (Part III, section 3). It is now time to turn to the objects of emotions.

Section 4: Objects as serious concern-based construals

In this section I argue that emotions are *typically caused* by concern-based construals, and *necessarily directed* at such construals. In the next section I will argue that, when understood in the context of a revised version of the Relational Schema, concern-based construals are also *sufficient* for individuating emotions.

1. *Are construals necessary?*

So far I have talked of emotions' *causes* as serious concern-based construals. The first question is whether the claim is that emotions are *necessarily* caused by concern-based construals⁵⁵.

Suppose my anger at you is caused by my having overdosed on coffee. The example can be interpreted in two ways: either the coffee caused/ enhanced my *construal* of you-as-offender or it *directly and sufficiently* caused the emotion. In the former case the emotion is still caused by a construal, so this should not be a problem for my model. Assuming that the second case is possible, it seems to generate a dilemma:

Either (a) what the coffee has caused is a genuine emotion, in which case construals are not necessary causes of emotions;

or (b) my putative anger is not an emotion but something weaker, like a mood⁵⁶.

The first horn threatens my model with including *too much* in the category of emotions: concern-based construals were meant to provide the criterion for discriminating emotions from other mental states, and if they are not necessary causes then they cannot constitute such a criterion. For example, a range of pathologies – like missing one of my prefrontal lobes and being either perpetually jubilant or depressed – would count as an emotion.

The second horn crystallises the opposite problem – the model's not capturing *all* the phenomena that seem to be emotions. When I am angry because (causal) of having had too much coffee, I seem to be in the grip of genuine anger *at you* and as far as I am aware I have very good reasons for it: a week ago you picked the most beautiful flower in my garden⁵⁷. Calling this a non-emotion, then, would subvert the account's claim to capturing emotions' complexity.

The dilemma, I submit, arises from considering the relationship between an emotion and its cause in isolation from the other elements in the Relational Schema. Recall that the schema casts emotions as relations between subject, target, focal properties, cause, motivating

⁵⁵ I am indebted to Thad Metz for asking this question after my presentation of a paper on concern-based construals (Philosophy Spring Colloquium, Rhodes University, 2002). Thanks to all the participants in the discussion for helpful comments which resulted in a much tighter model of emotions than the one I presented to them.

⁵⁶ I am using 'mood' here rather vaguely. The intuition is the old-fashioned one that moods differ from emotions in not being directed at an object (Kenny, 1963; de Sousa, 1987). If the assumption is suspect, so much the better: the larger the category of emotions, the larger ultimately Theory's scope for ascriptions. R. J. Davidson offers a more interesting distinction between the two. Moods, he argues, modulate/ bias *cognition* while emotions bias *action* (in Ekman and Davidson, 1994: 52). On this account (b) will not present an obvious problem for me, though my claim that emotions are *necessarily* directed (at concern-based construals) will be a problem.

⁵⁷ Notice that this case does not reduce to the case where coffee caused me to construe you as a flower-snatcher. Rather, I was angry at you 'just like that' and my anger cast about for a reason. Fortunately, I remembered about the flower, and then had a reason. But the coffee is still the direct and sufficient cause of the anger.

aspect, aim, and propositional object. Now, typically, the cause of an emotion coincides with the target's focal properties, which (if they are actual) are the motivating aspect. These elements, then are best understood in terms of concern-based construals, too. Moreover, since the propositional object captures the agent's reasons for the emotion, it too is in terms of a concern-based construal. My suggestion, then, is that

an emotion is a mental state which is typically caused by a concern-based construal and necessarily directed at such a construal, where the directedness is understood in terms of the Relational Schema.

So even if my current state is caused by too much coffee, as long as it is directed at you-the-flower-snatcher, the focal properties of the target (you) are a concern-based construal and hence my state is an emotion. And, obviously, the construal involved can range from very weak ("you just *look* like a flower snatcher") to strong doxastic commitment ("I believe/ know you are one")⁵⁸.

The proposal avoids both horns of the dilemma. The second, recall, was the charge that my model did not allow enough phenomena to count as emotions. On the current suggestion, because our criterion for emotionhood is not solely the cause, we need not conclude that what I am experiencing is a mood as soon as the cause does not rationalise the emotion⁵⁹.

The first horn was a problem, because it implied that the account allows for certain pathologies to count as emotions. The requirement that a concern-based construal is necessarily the object of emotion eliminates the possibility of my perpetual and directionless joy to count as an emotion.

2. Are construals sufficient?

It may be urged, however, that the proposal will not eliminate other, object-directed, pathologies, like phobias or neuroses. For example, suppose that a certain woman is continually and intensely terrified by men⁶⁰. All of her friends make it their task for a year to show her a good, innocuous, generous man every day. At the end of the year she still fears men as much as ever. We would want to say, the envisaged objection goes, that her fear is not an

⁵⁸ I suspect that the stronger the doxastic commitment, the more likely that the cause coincides with the focal property; though, of course, there is no correlation between the strength of the commitment and the rationality of the emotion, since strong doxastic commitment does not guarantee the veracity of the content.

⁵⁹ Notice, this breaches de Sousa's Intelligibility Condition in the sense that there is no motivating aspect to my emotion. It follows that the emotion is inappropriate. It does not follow that it is no emotion at all, though, because the target's focal properties are real: you really did steal my prettiest flower; though of course if I had not overdosed on coffee I would have never been angry with you, and your theft is an *excuse* for, rather than a *justification* of, the emotion.

⁶⁰ I am indebted to Marius Vermaak for this example and for his suggestion that it may be a problem for my account.

emotion but some (albeit affective) pathological state or disposition. And the problem is that my thesis forces me to call this an emotion, since the woman is concern-based-construing men as dangerous (the concerns in question would be aversion to being harmed, the desire to lead a fulfilled, unencumbered life, and so on). The objection, then, is that even if concern-based construals are the necessary objects of emotions, they are not sufficient to distinguish emotions from other states. So this is a good time to explore the scope of construals.

The first thing that should be noted about this example is that it would only impugn the sufficiency of construals if the woman's state was indeed not an emotion. My opponent, I contend, is mistaken in denying that the woman's fear is an emotion. For the denial to work, he needs to persuade us that there is a difference *in kind* between the woman's fear and typical fear. Now, there can be two reasons, as far as I can see, why someone would maintain that object-directed pathological states are not emotions⁶¹:

(a) because they may involve endocrine or nervous disorders and hence have exclusively or mostly *non-rational* causes;

(b) and/ or because they are utterly *irrational*.

The first option does not make for a difference in kind between emotions and object-directed pathologies. As soon as we acknowledge that we are embodied agents (let alone acknowledging that we are purely physical systems⁶²), it becomes obvious that even the most kosher of our emotions will be to some extent dependent on our hormonal and synaptic idiosyncrasies. But as soon as this is acknowledged, the difference between an emotion embedded in a nervously/ hormonally 'normal' person and one embedded in a not so 'normal' person becomes one of *degree, not of kind*. Of course, we do want to draw the line at drooling, but very few of the droolers' emotions are directed at objects in the world, still less at their specific properties which rationalise the emotion. After all, if our men-fearing woman did not have much contact with real-life men, and was brought up on a combination of violent thrillers and schmaltzy dramas, then she is quite right to fear men.

Picking up on droolers, the second way (from *irrationality*) of impugning the emotional status of object-directed pathological states seems more promising. But the option raises the

⁶¹ Such claims are certainly not rare, though they are seldom justified. For example, Peter Goldie, despite his rich conception of emotions, informs us glibly in a footnote: "Phobias, like claustrophobia and other phobic fears...are neither character traits nor emotions. A phobic fear, such as a phobia of dogs, is a disposition to respond and act in certain ways on sight of a dog" (2000:13). The reason, he tentatively implies, is that such fears are resistant to change. I take it that this is a variant of the second reason (irrationality) for considering these states non-emotions that I discuss.

⁶² I do not urge this line of thought too forcefully, because full-blown physicalism will make talk of emotions' *objects* nonsensical. I owe this insight to Walter Brown.

issue of what counts as irrational and how irrational an affective state must be before it ceases to qualify as an emotion. Dillon has distinguished three ways in which an emotions can be irrational:

- (a) irrationality by virtue of “some fault of the reasoning process” by which we have arrived at the propositional thought involved in the emotion (1997: 237);
- (b) irrationality “located within the first-order emotions. They contain cognitions that are false or at least unwarranted, that fail...to track reality” (ibid.);
- (c) irrationality, where neither false nor irrationally formed belief is necessary, but the emotion itself is “unresponsive to reason”, (1997: 237).

Obviously Dillon does not think these irrationalities a reason to deny the emotional status of a mental state. But, for present purposes, simply agreeing with her would beg the question.

So let me start by noting that our woman definitely fulfills b) and c) and possibly a). In what follows I argue that, nonetheless, the difference between her fear and typical fear is one of degree and any temptation to cast it as a difference in kind stems from an overcommitment to rationality (contrasted with *irrationality*) as constitutive of emotions. Now obviously I would beg the question here if I reiterated that the woman’s is an emotion because she has concern-based-construed men as dangerous. And, since the example is rather contrived and minimal, it will not allow for exploiting intuitions which do not rest on the construal thesis. So I turn to Shakespeare’s *Winter’s Tale* and argue that Leontes’s utterly irrational jealousy differs only in degree from Othello’s (what I dubbed) counterfactually appropriate jealousy. My argument unfolds as follows:

- a. In order to show that Leontes’s jealousy is relevantly similar to the woman’s phobia, I argue that it appears to be, for all intents and purposes, a bout of madness or illness. I found the claim on two features of his jealousy:
 - (i) its sudden appearance and disappearance;
 - (ii) its neurotic character which is evinced in its utter unresponsiveness to reasons, his soliloquies, and in the way others treat him.
- b. For all that, the difference between his jealousy and Othello’s is one of degree, particularly of degree of rationality. Two similarities between Othello and Leontes are adduced:
 - (i) their actions are similarly passionate, rash, and soon lamented;
 - (ii) both are haunted by the same concerns and describe their state (in soliloquies and to others) in very similar terms, which evinces the phenomenological similarity of their states and hence strongly suggests that if Othello’s is an emotion, so is Leontes’s .

The Winter's Tale opens with two lords extolling Leontes's and Polixenes's long-lasting and faithful friendship. By the end of the next scene Leontes has commissioned Camillo to murder Polixenes. The reason is jealousy. And it is excusable if the reason for *that* is not too obvious from the following story: Polixenes is about to end his nine-month visit to Leontes's kingdom. Leontes urges him repeatedly to remain for another week. Polixenes insists on leaving. Leontes asks Hermione - his wife - to plead with Polixenes. She does so - in front of Leontes - and succeeds. Suddenly, Leontes wonders why her appeals succeeded and his failed, begins to suspect illicit attachments, and in no more than a hundred lines has worked himself up into raging jealousy.

In the scenes to follow, Leontes decides that the child Hermione carries is Polixenes's; imprisons Hermione; sentences his new-born daughter to death by exposure; tries Hermione and finds her guilty, despite her ardent defense by the entire court and by Apollo's oracle to boot. He surfaces from this nightmare when, as soon as he disregards the oracle's verdict, a messenger announces the sudden (and clearly fateful) death of his son. For the next sixteen years Leontes repents the loss of his children, wife, and friendship with Polixenes.

a. *Leontes's jealousy as a bout of illness or madness*

My claim that Leontes's jealousy is *unfounded* hardly needs defence after this story. The question is whether it amounts to something near-pathological.

(i) I think its sudden and arbitrary appearance and disappearance partially support this suggestion. Compare *Othello*. Despite the equal violence of his jealousy, we are deliberately led through every stage of the jealousy-forming process and are made to understand that Iago's 'evidence' would move even a less sensitive man than *Othello*. By contrast, *The Winter's Tale* plunges us into the furious jealousy of a man whom we do not know, and whose jealousy makes little sense not only to us but to every single other character in the play. The fact that it makes no sense to the other protagonists, least of all to his wife, combined with the suddenness, suggests that the jealousy is out of character⁶³. So we have Dillon's first and second conditions for irrationality satisfied (weird acquisition of the belief, and the content not tracking reality).

(ii) More direct support for the illness/madness hypothesis is provided by the state's utter *lack of responsiveness to reasons and evidence* (Dillon's third condition). Everyone in the play, especially Leontes's most trusted attendants, pleads Hermione's innocence. Leontes not only disregards their pleas but uses them to batten his jealousy. For instance, when Camillo

⁶³ Though the play indicates that selfishness and tyranny, by contrast, are not out of character.

observes that there is nothing harmful about Polixenes's staying because of the queen's entreaties, Leontes comments in a paranoid aside: "They're here with me already; whisp'ring" (1.2.217).

Again, when Camillo tells Leontes that Polixenes stayed "To satisfy your highness, and the [queen's] entreaties" (1.2.231), Leontes hysterically harps on the verb until the mantra itself becomes proof of infidelity:

Satisfy?
Th'entreaties of your mistress? Satisfy?
Let that suffice (1.2.232-5).

The obviously *neurotic* element in Leontes's jealousy, which these passages crystallise, is evinced throughout his dealings with his pleading retinue⁶⁴. Our sense of neurosis culminates during the trial, when Leontes repudiates the oracle's verdict in a confusion of tyranny, childish helplessness, and peevishness (3.2.140-1).

Finally, the characters themselves treat Leontes as the victim of illness and madness, while he is in the grip of jealousy. Camillo urges him to "be *cur'd*/ Of this *diseas'd* opinion" (1.2.296-7, my italics), and - warning Polixenes of the intended murder - invokes disease:

There is a *sickness*
Which puts some of us in distemper, but
I cannot name the *disease*, and it is caught
Of you that yet are well (1.2.384-7, my italics).

Similarly, Paulina repeatedly treats Leontes's jealousy as a disease: she announces that she has come "to *purge* him of that *humour*/ That presses him from sleep" (2.3.38-9, my italics), and calls herself his "physician" (2.3.54).

There is also the suggestion that the jealousy is no more than a bout of madness. Paulina accuses Leontes of "Not [being] able to produce more accusation/ Than your own weak-hing'd fancy" (2.3.117-8). And Hermione adumbrates the charge of madness during the trial: "My life stands in the level of your dreams" (3.2.81). Finally, Leontes himself retrospectively admits to having been "*transported* by my jealousies" (3.2.158, my italics) and the only objective voice in the play, Time, confirms the diagnosis by referring to Leontes as subject to "his *fond* jealousies" (4.1.18, my italics). Leontes's phrase suggests madness, Time's - delusion.

The madness/ illness hypothesis is important for three reasons. First, it shows Leontes to satisfy all three of Dillon's conditions for irrationality. Second, the fact that it is a *bout* of

⁶⁴ When Paulina comes to plead with him, his new-born daughter in her arms, and all the lords support her, Leontes repeatedly accuses them of being traitors and her of being unnatural (2.3.47-192).

madness or illness suggests that Leontes is overall a rational agent and this is a slip in his rationality, rather than characteristic (just as our woman is, presumably, overall a rational agent). Third, it suggests that even if the jealousy is non-rationally acquired (like a disease or madness is), we are still entitled to - and the characters in the play do - censure Leontes for the emotion's unresponsiveness to reasons *subsequently* to its acquisition. Of course, the disease and madness imagery may be no more than tropes. I think, however, that chiming with the other two features of Leontes's jealousy, it serves to confer at least as much of the neurotic, mad and sick on Leontes's jealousy as there is in the example of the woman who fears men (though of course in her case it is a more permanent disposition). If this is not granted and it is urged that we normally use such language to characterise the more disturbing of our emotions as well as our pathologies in general, this lends further plausibility to my claim that certain pathologies involve genuine emotions.

b. *The difference between Leontes's and Othello's jealousies is one of degree*

Hopefully I have offered enough evidence that Leontes's jealousy has all the appearances of a bout of illness, madness, or flash-neurosis. The next task is to show that, despite appearances, Leontes's jealousy is an emotion which differs only *in degree* (of rationality), rather than in kind, from Othello's.

(i) The first similarity between the two protagonists is that they both act hastily and vengefully towards their wives and rivals, and soon afterwards regret it. Othello commissions Iago to murder Cassio and himself murders Desdemona; Leontes commissions Camillo to murder Polixenes and condemns to death Hermione at a pseudo-trial. Similarly, both repent their actions as soon as performed. In response to Emilia's revelation of the truth about the handkerchief, Othello berates himself with "O fool! fool! fool!" (5.2.326), for having believed Iago. Leontes, sobering at his son's death, laments having "too much believed mine own suspicion" (3.2.151) and letting himself be "transported by my jealousies" (3.2.158). In response to Paulina's apology for her harshness, he cries "I have deserv'd/ All tongues to talk their bitt' rest" (3.2.215). By analogy with Othello's suicide, Leontes cloisters himself until his wife and daughter are discovered living.

It may be objected that all these actions are perfectly compatible with a range of other, unemotional, states. For example, a psychopath can murder his wife and her lover (assuming she has survived long enough to have one) without being motivated by jealousy. The objection overlooks two things. First, we have agreed that Othello's is an emotion, and all I am doing is showing that Leontes's actions are very similar to Othello's. This confers *prima facie* plausibility to the claim that the two men's emotions differ in degree only.

But we can do better than that. Recall (and this is the second point) that a *constitutive* part of the emotion, according to the Relational Schema, is the emotion's *aim* which is defined in terms of characteristic expressive behaviour. And recall too that what is characteristic will be constrained by the biological as well as social functions and origins of an emotion. Now, murdering one's wife and rival, though legally unacceptable in Western society, is certainly a typical expression of jealousy throughout the ages and some other cultures today (it is even an *expected* expression – consider Shariah law).

Of course, murder and violence are *actions*, rather than expressive behaviour. But we also acknowledge that emotions have typical ensuing actions. If we did not do that, emotions would not be a useful explanatory tool for action (which they are⁶⁵, *pace* Griffiths, 1997). Furthermore, and this is de Sousa's notion of aim, in both cases violence to others is not the only expression of the emotion: both Leontes and Othello are noticed (by their wives) to have a "brow of much distraction" (*WT*, 1.2.149); both are impatient and suspicious of those around them; both share (equally persuasively) their torment with the audience. I will come to the last point in a moment. The present thought is that, though not sufficient, the aim and actions are partly *constitutive* of the emotion and so certain behaviour (when sincere) compels us to consider its originating state an emotion (more of this in Part III, section 1).

But suppose it is objected that the reason Leontes's and Othello's actions are so similar is that Leontes's state is not sufficiently madness/neurosis/illness-like, and so I have subverted my earlier labours to show the contrary. I will not be moved by this until I am shown the difference between our neurotic woman's cringe of fear when a man walks into the room and non-neurotic cringes of fear. There is not, in other words, a significant difference between behaviour out of a violent emotion and behaviour out of a pathology. Fine, my opponent pursues remorselessly, then it is question-begging to use behaviour as the fulcrum of an argument for integrating pathologies into emotions. Fair enough, but we are on the same ground here. The only way to avoid a *tu quoque* is to adduce further reasons for the similarity of Othello's and Leontes's states. Then considerations of behaviour will carry legitimate weight, since my ultimate claim is that the elements of a revised Relational Schema – one of which is the aim – are jointly sufficient for a state to count as an emotion.

⁶⁵ I appreciate that I am merely stipulating emotions' unity as an explanatory category. Two considerations help make the claim a little less facile. First, we appeal frequently to emotions in explaining each other's actions (more of this in Part III, section 1 below). Second, my overall argument assumes that we appeal to distinct categories of mental states in our attributions and predictions. If I am required to show *that*, I would be embarking on a new project (and possibly a Ph.D. dissertation). So I have to assume, at least to an extent, that emotions are a useful explanatory category (cf. fn. 36, p. 40).

(ii) That Othello's and Leontes's states are similar is obvious from their soliloquies.

First, once they suspect their wives, sleep deserts both. Second, their suspicions are refracted into utter disillusionment with women's honesty in general. Othello suddenly heeds Iago's warnings about the promiscuity of Venetian women and Desdemona's deception of her father; and becomes obsessed with cuckolds. Similarly, Leontes devotes an entire speech to the topic (1.2.190-207). Here is an excerpt:

There have been... cuckolds ere now,
And many a man there is ...holds his wife by th'arm,
That little thinks she has been sluic'd in 's absence,
And his pond fish'd by his next neighbour.

The image, with its simultaneous vividness and vulgarity, suggests a painful revulsion with women. The fact that the speech is addressed to his young son, hardly an apt audience, evinces the profundity of the disillusionment and the ensuing loneliness.

Finally, and most significantly, once the suspicion is lodged, neither Othello nor Leontes thinks of anything else, and both interpret every event around them as an omen of their wives' infidelity. When Leontes himself sends Hermione and Polixenes for a walk in the garden and they go, he exclaims woundedly "Gone already!" (1.2.185). When Camillo defends their innocence, Leontes urges hysterically:

Is whispering nothing?
Is leaning cheek to cheek? Is meeting noses?
Kissing with inside lip? Stopping the career
Of laughter with a sigh?...
... Wishing clocks more swift? (1.2.284-9).

Significantly, he moves from the act which he could have witnessed and possibly did (whispering), through acts which he could not have seen because not performed, to finally knowing the actors' wishes. The progression captures a crucial feature of jealousy in general – the way it thrives on itself; a feature familiar to Iago and exploited by him after a glib aside to the audience:

Trifles light as air
Are to the jealous confirmations strong
As proofs of holy writ (*O*, 3.3.326-28).

We have, then, at least two good reasons to consider Othello's and Leontes's jealousies different in degree only: both protagonists' actions and expressive behaviour are similar, and moreover typical of jealousy; both describe their own states in very similar ways, and moreover the descriptions fit traditional notions of the phenomenology of jealousy. I suppose the move from similarity of description to similarity of phenomenology might be resisted on the grounds that both plays were written by the same author, and so it is natural that the protagonists'

concerns and language are similar. The objector forgets Shakespeare's proverbial prolific vocabulary and diversity of characterisation. Considering that, it seems unlikely that he would use such similar concerns and language in two different plays, unless he thought that the two phenomena are analogous, and are instances of the same emotion, let alone of emotion in general. This is no appeal to authority⁶⁶: as I argued at the outset, the reason we consider Shakespeare the *Maître* is that we are moved by his plays. That in turn indicates that he captures something real about human situations in a convincing manner. So, in the *Leontes* and *Othello* cases we are genuinely moved (which means Shakespeare has captured something real about jealousy in both cases); and are moved in similar ways by the protagonists' emotions, despite their different circumstances (which means that the difference is one of degree rather than of kind).

If the second half (b.) of my argument worked, we should have no trouble concluding that the difference between *Leontes*'s and *Othello*'s jealousies is one of degree, and so *Leontes*'s jealousy is an emotion after all. True, he arrives at the concern-based construal at which it is directed in a much more irrational (possible non-rational) manner than *Othello*. But this is no reason to deny *Leontes*'s jealousy the status of an emotion. If that is right, and the first half (a.) of my argument worked – that is, *Leontes*'s jealousy is relevantly similar to our men-haunted woman's fear – then it would follow that certain neuroses and general object-directed pathologies similarly involve genuine emotions. This seems right not least because for a long time emotions themselves were treated as pathologies. I take it that this point supports my thesis: emotions have had their bad press from their more violent and irrational exemplars. Interestingly, it is precisely these exemplars which seem most pathological, since they are unresponsive to reasons and evidence. But the folk temptation to lump other emotions into the pathology category evinces the sound intuition that we are talking about a single category after all. My proposal is the obverse of the folk one: instead of working from irrational emotions to the conclusion that all emotions are pathologies (because beyond the pale of rationality), it would be more fruitful to start with the rational emotions and then notice that object-directed affective states which seem like pathologies are merely the slovenly cousins – but cousins nonetheless – of rational emotions.

Conclusion

Concern-based construals, then, are

⁶⁶ I am indebted to Tom Martin for pointing out that this argument may be read as an appeal to authority.

- (a) the typical causes and necessary objects of emotions;
- (b) potentially sufficient to confer emotion-status on a state, when understood in the context of the Relational Schema.

This is not the complete picture yet, though, which is why we need the “potentially” above. Two aspects of my discussion of pathologies are noteworthy. First, my argument invoked rather dodgily the notions of aim, which is in the schema but is unexplained; and of phenomenology, which is not in the schema (yet). Second, I employed Dillon’s criteria of irrationality, when I already set up (section 2) very different ones. In the next part I remedy these deficiencies.

Part III: Emotions as attitudes

Introduction

In this part I explore emotions *qua* attitudes in an attempt to gather all the loose ends from the previous part. In section 1 I return to emotions’ aims and to the use I made of them in the last section. In section 2 I argue for an addition to the RS - phenomenology. In section 3 I consider, for the first time systematically, rationality; and argue, together with Goldie, that the notion is pretty useless in the context of emotions. Instead, I shift to talk of intelligibility and appropriateness by extending Goldie’s notion of narrative which provides (a) the basis for individual notions of emotional appropriateness; and (b) the locus for explaining construals and tightening the Relational Schema. In section 4 I formalise the notion of narrative by casting it in terms of a Narrative Relational Schema.

Section 1: Aims

Recall that one of my arguments for the difference in degree between Othello’s and Leontes’s emotions turned on the similarity of their behaviour and the behaviour’s in turn fulfilling the formal aim of jealousy. Under the rubric of behaviour I included immediate expressive behaviour, like a brow of much distraction; actions, like killing one’s wife; and border-line cases, like being in general suspicious and impatient. Two things are to be noted about this rather ambitious inclusion.

First, the notion of aim is the only element of the Schema which characterises an emotion as an *attitude* rather than the emotional *object*. This may seem rather out of place, and so it would be if RS formalised a relation between the subject and the object of the emotion. Recall, however, that it was supposed to characterise *fully* an emotion, which means that what

sort of attitudes emotions are will figure crucially in our account. We are embarking, then, on an exploration of emotions as attitudes.

Second, it is unclear whether I have diverged from de Sousa's notion of aim. Recall that the aim was a constituent of RS_i and was defined in terms of the immediate expressive behaviour of the emotion. Now, this seems to encompass purely involuntary behaviour (like Ekman's facial expressions, and Jamesian bodily attitudes which are typical of emotions) on the one hand, and more action-like behaviour on the other. So de Sousa writes:

the *motivational role* of emotions defines their characteristic aims and acts as a constraint on the character of each specific emotion (1987: 121, my italics).

Presumably, the notion of motivation implies *action* rather than mere expressive behaviour. The examples that de Sousa offers, however, are in the grey area between action and mere behaviour: paradigmatically, we jump from joy, slouch with depression, etc. (1987: 121). So, the question is how much our category of aim should include. David Velleman's position is illuminating in this respect. He argues, in a different context, that the distinction between "mere happenings" and "autonomous action" (2000: 4) does not exhaust behaviour. He introduces a third, in-between, "category of ungoverned activities", and defines them as

the things one does rather than merely undergoes, but that somehow one fails to regulate in the manner that separates autonomous human action from merely motivated activity (2000: 4).

He dubs this category "mere activities". Paradigmatic cases are activities like tapping one's foot, braking to avoid an accident when thinking about something unrelated to driving, and generally things that we do unconsciously or not deliberately but *do* nonetheless.

I suggest, then, that de Sousa's aim be defined in terms of expressive behaviour, understood both as mere happenings (like racing pulse, schadenfreude grin, and generally uncontrollable bodily expressions) and as mere activities. Actions, on the other hand, will not be part of the schema, and hence not part of the emotion, though they will be part of the narrative episode in which the emotion is embedded (cf. section 3 below).

Section 2: Feeling towards

It should be obvious that, despite de Sousa's elaborate analysis of the emotional object, tucking on some expressive behaviour, will not furnish the Schema with exhaustive insight into what sort of attitudes emotions are. Of course, the object is crucial, since there is a constitutive relation between an attitude and its formal object (e.g. belief is partially defined in terms of truth). But it still seems that the Schema has omitted the fulcral personal element about

emotions. And I do not mean the personal *perspective*, which is captured by the central causal and reason-explanatory role concern-based construals play. Rather, the problem is that the Schema ignores the *phenomenology* of emotions. Such phenomenology would not form part of a relational schema capturing propositional attitudes, because it does not feel like anything to believe that Paris is the capital of France, and most of the time desires do not have a particular phenomenology (unless they are emotional desires). By contrast, we intuitively take the phenomenology of a particular emotion to be definitive of the emotion.

I suggest that Goldie's notion of "feeling towards" is most felicitous for capturing this phenomenological element. He defines it as "thinking of with feeling, so that your emotional feelings are directed towards the object of your thought" (2000:19). Two things are to be noted about this definition. First, the notion is not, and certainly Goldie does not intend it as, a reductionist notion in the sense of being constituted by a thought *plus* a feeling. Second, it does not help invoking "*emotional feeling*" when we are trying to capture what it is about feeling that is partially definitive of emotion as an attitude. I think that the conception can be strengthened enormously, and the two worries assuaged, by the idea of serious concern-based construals.

Goldie argues for the *sui generis* nature of "feeling towards" on the basis that thought *simpliciter* and *feeling-laden* thought have different contents (2000: 60). He is a little vague and inconsistent, though, as to what precisely that means. His intuition is that "feeling towards an object...is feeling towards that thing as having certain properties or features" (2000: 54). This sounds very much like the construal thesis. But Goldie erodes this conception of feeling towards when he continues: "thinking of and grasping the saliences of a thing is one matter and having feeling towards that thing is another" (2000: 59). When he tries to specify the difference, though, he returns to construals (though he does not use the term): the difference, he claims, is partly a phenomenological one, and partly in "*the way* of grasping the saliences of the object of the emotional experience" (2000: 59, *my italics*).

I think to avoid inconsistency here we should abandon Goldie's insistence on the distinction's being in terms of content. It is quite obvious that my thought about something can equally have as its object a construal (recall Roberts's claim that all mental activity consists of construals). Moreover, presumably we can have thoughts about our concern-based construals: so, when Roberts was relating his ketchup episode, the content of his thoughts was his concern-based construal of his daughter as guilty. What was missing in the *thought* about this construal was, I submit:

- (a) the *seriousness* with which he was thinking about it during the episode;
- (b) (hopefully) the expressive *behaviour*;

(c) the distinctive *phenomenology* of anger.

The distinction, then, between thinking of and feeling towards becomes (as is obvious from the terms) a distinction in *attitude*. In feeling towards, but not in thinking of, the concern-based construal is (a) taken seriously; (b) gives rise to expressive behaviour (both mere happenings and mere activities); and (c) has a phenomenological feel, where phenomenology typically includes proprioception and a general sense of one's state of mind (normally the locus of this sense would be a sense of the personal significance of the construal in question).

Emotions, then, are attitudes characterised in terms of the revised Relational Schema:

RS*: R (*S t f a c p m s*),

where **R** is the emotion type, **S** is the subject, **t** the target, **f** the focal property, **a** the motivating aspect, **c** the cause, **p** the propositional object, **m** the aim, **s** the characteristic feeling towards the object of the emotion (including seriousness towards the construal and characteristic phenomenology); and where the focus, motivating aspect, and (if any) propositional object are *necessarily* a concern-based construal, while the cause is only *typically* a concern-based construal.

RS* is an improvement on all of the three writers on whom it draws:

1. It is a refinement of Goldie's notion of 'feeling towards' in that it distinguishes it from thought in terms of *attitude* rather than content; and embeds it in RS* where it is no longer a vague notion;

2. It is a refinement of Roberts in that he ambiguated between construals as attitudes and construals as objects. If they were attitudes the problem of their causes arose (section 4); if they were objects, they were not useful enough without the complex picture both RS and RS* provide.

3. RS* refines de Sousa's RS by offering a richer account of both the object of emotion (by casting its various aspects as construals) and the sort of attitude emotions are (by taking cognisance phenomenology and seriousness);

Section 3: Rationality

Recall that RS provided criteria for intelligibility and appropriateness, criteria which I did not use in my discussion of the men-haunted woman and Leontes, in order not to beg the question. It is time to return to the sore topic of rationality. It should be obvious by Dillon's discrimination of the three ways in which an emotion can be irrational, that talk of rationality is not particularly helpful *vis-à-vis* emotions. For one, the constitutive role of construals precludes our emotions from tracking reality perfectly most of the time (Dillon's second criterion of irrationality). Second, the notion of construals renders otiose the idea that beliefs, or any reasoning process, are necessarily involved in emotions. So talk of the rationality of the reasoning process behind emotions (Dillon's first way emotions can be irrational) is not helpful either. Finally, responsiveness to reason *does* matter, but it is not clear that it is the apt criterion for irrationality rather than for something like inappropriateness. For example, suppose that I am desperately in love with a real bastard of a man. All my friends try to dissuade me from continuing the relationship on the basis that he has abused violently his first ten wives, that he is insensitive, and whatever else friends say in such cases. Is my emotion irrational? Well, in one sense (and certainly in Dillon's), yes. But perhaps being with this man is *strategically*⁶⁷ very rational for me: the man is partly definitive of my femininity in that he is such a brute of a MAN; or I perceive the world as hostile and, feeling vulnerable, prefer to be abused by one person who defends me against the rest of the world's brutalities; or whatever.

1. *Emotional appropriateness*

The discrepancy between what is *strategically* rational and what is *cognitively* rational in this case is enough to cast serious doubts over the usefulness of the notion of rationality in the context of emotions. So, de Sousa, for example, talks of a third *sui generis* type of rationality, "axiological rationality" (1987: 171) which is applicable to emotions only. My suspicion is that, despite our everyday practice of accusing emoting (female) people of irrationality, the notion is pretty useless. Peter Goldie has argued, instead, that notions of intelligibility, appropriateness and proportionality are much more apt here. The notions (only two of which I discuss) are, I contend, isomorphic with both de Sousa's axiological rationality and with the conception of rationality that the RS* naturally generates. So let us see what that is.

Recall that according to de Sousa's RS, an emotion is *intelligible* if:

- (a) its causal explanation coincides with its reason explanation (that is, cause coincides with focal properties); (b) and *as far as the agent is concerned*, the focal property is an instantiation of the formal object of the emotion.

⁶⁷ I owe the distinction between cognitive and strategic rationality to de Sousa (1987: 163-4)

An emotion is appropriate if

(a) it is intelligible;

(b) and “the target actually has a focal property in virtue of which the formal object fits the target”.

The intelligibility conditions seem fine as they stand, as long as causal and reason explanations in (a), and the phrase “as far as the agent is concerned” in (b) are all understood in terms of concern-based construals rather than in terms of beliefs or judgements. As far as appropriateness is concerned, however, this cannot be the whole story. Considering the central role concern-based construals play in RS*, our notion of appropriateness must be more fine-grained.

Before I develop this point, two clarifications are necessary. First⁶⁸, what is at stake here is *not* moral appropriateness. I am concerned with a notion of appropriateness internal to the conception of emotions developed here. An analogy with doxastic appropriateness would illuminate this idea of internal appropriateness. When we judge a 21C person’s belief in witches inappropriate, the judgement is not a moral one. Rather, we are censuring the subject for having violated a norm internal to the notion of beliefs. Of course, if this person’s belief led to a witch hunt, we would judge it *morally* inappropriate. But the judgement is not internal to the conception of doxastic appropriateness. Rather, it is a judgement of a morally inappropriate *link* between his *belief* and the ensuing *action*. When I talk of emotional in/appropriateness, then, I mean violation/observation of the norms of appropriateness entailed by the concept of a particular emotion (as captured by RS*). In this sense an emotion can be emotionally inappropriate and morally appropriate, and vice versa, just as a belief can be doxastically inappropriate but morally appropriate (for example, the belief that I should save that drowning child is morally appropriate, but would be also doxastically inappropriate if its basis-belief, that there is a child drowning, was false or irrationally formed).

Secondly, it should be noted that both de Sousa and Goldie, when discussing appropriateness, unwittingly, I hope, oscillate between an objective notion of appropriateness and a notion of what we (whoever we are) would judge appropriate in any given situation. As soon as the distinction is drawn, it should be clear that the first would be something like rationality, and that we want to focus on the second.

But as soon as that is noticed, condition (b) must be abandoned or at least modified, since a lot of the time we (the same ‘we’ making the judgement of whether an emotion is

⁶⁸ I am indebted to Ward Jones and Eusebius McKaiser for making me take seriously the danger of being misconstrued here, and hence of the argument’s not going through. Hans-Johann Glock first objected to my notion of appropriateness from the viewpoint of moral appropriateness.

appropriate) do not know whether the target has the construed property in question. For example, if condition (b) is in place we could not judge someone's fear of Hussein's nuclear weapons appropriate or inappropriate: we just do not know whether he has such weapons or not. What we do in practice, and do well, is to go with certain reports of whether he has or not, reports that are a matter of fashion, or world politics, or whatever other cultural and historical whims are in place; and then judge *on the basis of that* whether a certain person's fear of Hussein's nuclear weapons is appropriate. This point would be even more pertinent for all forward-looking emotions (hope, for example) which are not caused by, or directed at, an object's or situation's current properties but rather whose object and cause are a future, desired construal of them. Yet we do judge people's hopes appropriate or inappropriate⁶⁹. For example, if an Oxford graduate who has been offered and *wants* a teaching job, told us that he ardently hopes that the Calcutta municipality will give him a job as a street-sweeper, we would raise our eye-brows and any story he tells us would at best make the hope intelligible. So condition (b) is in need of some serious revision.

2. Narratives

Recall that I elucidated concern-imbuement by Rorty's notion of magnetizing dispositions. Magnetizing dispositions, in turn, were shaped by the personal significance of the events in a person's life. This is precisely Goldie's notion of a narrative structure:

Our lives have a narrative structure - roughly speaking, they comprise an unfolding, structured sequence of actions, events, thoughts, and feelings, *related from the individual's point of view* (2000: 4, my italics).

The crucial point of similarity with Rorty is the emphasised phrase⁷⁰. Let me return to Tom Pinch. For over four hundred pages Dickens shamelessly sets him up, by gradually and constantly disclosing more and more of Pecksniff's nastiness, and simultaneously showing how Tom excuses away its every instance. Pecksniff is the locus of Tom's narrative, a sort of personal god, to whom all of Tom's thoughts about self and others infallibly defer. The element of divinity is crucially revealed in the penultimate sentence of the chapter from which I cited earlier:

Pecksniff had gone out of the world - had never been in it - and *it was as much as Tom could do to say his prayers without him* (32.433, my italics).

⁶⁹ Again, I do not mean here a judgement of *moral* appropriateness, which we also do. So, if I told you that I hope that all the grass on campus and the harvest of Southern Africa withers overnight, the inappropriateness here would be a moral one. If I also told you that there is no reason for my hoping this, the hope would be emotionally inappropriate.

⁷⁰ In an e-mail, Goldie writes that he has not laid sufficient emphasis on this point; and defines narrative, instead, as the subject's "representation" of these events.

Together with the passages adduced earlier, this episode evinces Tom's loss of his self-conception and his utter grief. Now, if we are to follow de Sousa's second criterion of appropriateness, we should conclude that Tom's grief is inappropriate. He mourns having lost his locus of identity, of ethos, of humanity. But Pecksniff should never have been such a locus, as Tom evidently and painfully realises. So, the target never had the focal properties mourned now and envisaged by his earlier love and respect. The notion of narrative, then, allows for Tom's emotion to be appropriate. It is only right that when the very root of our self-conception is eroded we should feel grief, and feel it utterly.

Furthermore, narratives illuminate the appropriateness of Tom's earlier love and awe of Pecksniff. Many of the characters in the novel work on de Sousa's notion of appropriateness and see Tom as a fool (if not an idiot) for loving and revering Pecksniff. The narrator, by contrast, urges the opposite judgement: it is Tom's character, his innate generosity, on the background of the other characters' utter selfishness, which makes for his favourable construal of Pecksniff, and which the novel celebrates⁷¹. Reading Tom on de Sousa's terms would preclude us from noticing this. Worse, it would make us judge appropriateness like the 'bad guys' do⁷².

This is not to claim a subjective status for appropriateness. On the contrary, the notion of narrative gives us the framework for talking about "paradigmatic narrative structures" (Goldie, 2000: 33). There are two senses of narrative structure here. The one concerns individual emotions: what caused the emotion, how and over what period did the emotion unfold⁷³, etc. Call this the *emotion's* narrative structure. The second is a *person's* narrative - the events in his life in their personal significance. And the point is *not* that as long as my current emotion *coheres* with my narrative, it is automatically appropriate. That would be to conflate appropriateness and intelligibility⁷⁴. Rather, the person's narrative is important for

⁷¹ That Dickens means Tom to embody a positive vision is obvious not only from the moving way in which he solicits the reader's sympathy for Tom, but also from his own admission in the "Preface To The First Cheap Edition" that the object of the novel is to explore "Selfishness" (ix).

⁷² The fact that the *good* guys judge the emotion appropriate does not make it *morally* appropriate. Rather, the distinction between the bad guys and the good guys is useful here because Dickens's bad guys normally also have a limited understanding of human nature. Their judging an emotion in/appropriate, then, is not a good indication of whether the emotion is indeed in/appropriate.

⁷³ I am assuming here that an emotion can be long-standing (my love for my mother), as well as an episodic state. Paul Ekman (1994) has argued that no state which lasts over a couple of minutes counts as an emotion. I think that apart from this claim's being counterintuitive, it explicitly jars with the assumption of this thesis that emotions are useful for explanations of actions and other mental states. If Ekman is right, then the assumption is deeply flawed: without recourse to emotions as long-lasting states or as processes, one cannot explain the bereft husband's behaviour by reference to his grief, one cannot explain Lear's gratitude to Cordelia, or his moral awakening before she dies, and so on.

⁷⁴ I am indebted to Hans-Johann Glock for this insight.

placing the emotion in the right context. Then it is the comparison of the emotion so placed to the emotion's *paradigmatic* narrative structure, that does the evaluation job. And so, considering an individual's narrative is important in working out the narrative structure of his emotion, and then comparing this structure to the paradigm. For example, the narrative palimpsest that most of the characters in *Martin Chuzzlewit* apply to Tom's narrative is a paradigmatic foolishness narrative. From this vantage point, Tom's love and later grief would seem intelligible but inappropriate, because love and grief are supposed to track to some extent the properties of the target. The narrative for Tom that I am urging is one whose warp is generosity and woof - self-conception. Considering this, it is quite appropriate to celebrate one's paragon of humanity and lament the death of one's self-conception.

Of course, a *society's* narrative and self-conception will also play a large role in determining a paradigmatic narrative structure for each individual emotion. But to concede this is to corroborate the insight that our notion of appropriateness must be cast in terms of narratives, rather than in terms of a one-to-one map between construals and their targets⁷⁵.

Section 4: Taking narratives seriously

To say all this, however, is only to confer some *prima facie* plausibility to the narrative notion, but not much clarity. Let me start with what we normally mean by narrative. Gregory Currie defines narrative in a broad sense as "the means by which a story is told" (1998: 654). This fits, rather metaphorically, with Goldie's notion of "an unfolding structured sequence of actions, events, thoughts, and feelings, related from the individual's point of view". But an immediate puzzle is conjured by Goldie's definition. Events and actions are of a different *kind* from thoughts and feelings. To pursue the narrative metaphor - for that is all it is at this stage - thoughts and feelings are already a sort of commentary on events in the world; and actions are both events and stem from the commentary on other events.

I think the puzzle is generated by lack of clarity about who the putative narrator of the narrative is. So let us pursue this track. Goldie makes it sound like it is the subject of the narrative. This cannot be right if the narrative is to capture the full causal story of the emotion, because the subject is often unaware of the causal interactions between a particular event-construal of his and his ensuing construals. If he was so aware, we would have infallible access

⁷⁵ A typical example of the necessity of paradigmatic narratives for judgements of appropriateness is the psychoanalyst's unearthing of a person's story. Only once the story emerges, can the psychoanalyst judge the current emotion appropriate or inappropriate. (I am grateful to Ward Jones for this example.) A more quotidian use we make of narratives is when we ask Sue to forgive Joe for having been rude, because 'he hasn't been himself since his wife died'. I argue in the **CODA** that this cannot excuse Joe's having *acted* rudely. But the point is that we cannot start to judge his *emotion* (anger or irritation at Sue) in/appropriate without knowing Joe's story.

to the contents of our minds, which is implausible. So, recall that Jonah was unaware that his construal of his boss as domineering stemmed from his formative family troubles. The family troubles in turn consisted of actual events and his construal of them in particular ways. If we return to Rorty's magnetizing dispositions, these were formed partly by actual (unconstrued) events in the world (like Jonah's having to stay with his grandfather); accurate construals of other events (like perhaps grandpa's references to his beloved daughter-in-law over some brandy); and finally, a purely illusory homicide-construal of scarf-tying. If Jonah were the narrator of this story, the narrative could be captured in the following Relational Schema:

NRS1: $R_c(cE1, \dots, cEn)$,

where R_c is the construed relation among the construed events $E1$ to En (where construals can be emotions, thoughts, and construed causal links) .

But, as I have already said, a construed relation of construed events cannot capture the whole story, because Jonah is aware neither of all his *construals* nor of their relevant and actual causal *interrelations*. So, a fuller conception of a narrative must include all the *actual* construals and their *actual* causal relations. The actual relations can be captured in the following schema:

NRS2: $R_a(cE1, \dots, cEp)$,

where R_a is the *actual* relation among the *construed* events $E1$ to Ep , and where Eo to Ep are the construed events of which Jonah is *unaware*.

So the temptation is to cast the relevant narrative structure as a second-order relation between NRS1 and NRS2. But this is too optimistic. There are events out there which have played their causal role in Jonah's life qua actual unconstrued events. So we would need to add that in. Moreover, if the narrative is to be complete, we need the actual events $E1$ to Er , too, to figure in it. This can be summed up as:

NRS3: $R_a(E1, \dots, Er)$,

where R_a is the actual relation among the actual events $E1$ to Er , and where events Eq to Er are the actual events which played their causal role in Jonah's life unconstrued.

Now it seems the appropriate narrative relational schema is in place:

NRS: $R[R_a(E1, \dots, Er); R_a(cE1, \dots, cEp); R_c(cE1, \dots, cEn)]$ ⁷⁶

So who is the narrator? The schema, I submit, allows for no one short of God or at the very least an archangel. The reason is that the schema captures an *actual* second-order relation which is not accessible to Jonah or to any other mortal. The schema is still narratorless, in other words, and so cannot be the schema of a narrative.

⁷⁶ The only reason the final permutation is missing (that is, the construed relation between actual events) is that, as I have argued, a subject never has unmediated access to the world.

An essential part of our narratives are our interpreters of these narratives⁷⁷. I mean 'essential' here partly in a causal sense - others perpetually comment on our character and our doxastic and emotional interactions with them; and so make us re-see and re-cast ourselves. But of course, another sense in which others are involved is normative: making sense of my own narrative involves norms of appropriateness which can shape my narrative in at least two ways:

- (a) if I am good, I would aim to square my attitudes and behaviour with the norms of appropriateness (like Tom Pinch does);
- (b) if I am bad, I would aim at *rationalising* my current attitudes and behaviour into *seeming* to fit the norms (like Jonah does).

This points to the *constitutive* role interpretation plays in my narrative. But as soon as we have invoked an interpreter, he too, whether he is an individual or a society would have his own narrative. So it may be thought that the second-order relation envisaged by the NRS becomes the monstrous third-order relation:

NRS*: R{ R[Ra(E1,...,Er); Ra(cE1,...,cEp); Rc(cE1,...,cEn)]; Rc*[Ra(E1,...,Er); Ra(cE1,...,cEp); Rc(cE1,...,cEn)]},

where Rc* is a relation as construed by the interpreter, among the elements of the other's narrative.

Or to make it look at least a little more endearing:

NRS*: R {NRS; Rc*NRS}, where NRS can characterise either a person's entire narrative or any episode of it (like an emotion).

Again, I submit, this schema is a divine one (though it certainly does not *look* it)⁷⁸. The problem is that it still captures an *actual* relation between the interpreter and the interpreted and there are elements in this relation which are available to neither. We need, then, to dispense with all actual relations, and replace God with a construing human being. We also need to take into account the fact that the interpreter does not have access to most of the other's construals of the relevant events, but does have access to other events which the interpreted lacks. The schema then becomes:

NRS: Rc*{1/x (cE1,...,cEn); (c*E1,...,c*Er)},**

where Rc* is the relation, as construed by the interpreter, between a portion (1/x) of the other's construals of events E1 to En, the interpreter's own construal of these events, and his construal of others which are unavailable to the interpreted (Eo to Er).

⁷⁷ I am grateful for this insight to Undine Weber.

⁷⁸ I am indebted to Paul Trompeter for this insight and a discussion of what follows.

Perhaps my schema is a little too unwieldy. But seeing that our mutual interpretations are so complex as a matter of fact, it is no surprise, and I hope no cause for embarrassment, that a relational schema which attempts to capture these relations should be so outlandishly cumbersome.

But it may be further objected that I am courting an infinite regress here: after all, the interpreter's construed relation of the other's construals seems remarkably like the interpreter's own narrative; and so it seems that the first half of the formula will have to be recast in terms of NRS** and so on ad nauseam. The objection becomes especially pressing when we notice that the subject of the narrative can also be its interpreter, i.e. fill out the Rc* relation. Fortunately, the regress is a mere illusion. For the regress to occur, the interpreter's own narrative must be *constitutive* of his construal of the interpretee. But this is not the case. The interpreter's narrative only figures *causally* in the construal of the relation: it *partially* biases the interpretation but does not constitute it. So the danger of regress is spurious.

It may, however, be further urged that the *normative* notion of an interpreter has no place in a *descriptive* schema. My intuition here is that a narrative is not merely a descriptive, but partly an interpretative and so normative, notion. And when we are dealing with its description we need norms to guarantee that what we are describing is a narrative rather than something else (like a history). For example, recall that in Part I, I said that certain normative commitments are partially constitutive of describing someone as believing that p: we would not describe his state as *belief* unless p figured at least some of the time in his inferences and he acted at least some of the time as if p was true. My suggestion, then, is that the interpreter - whether another person, or society, or self - is necessary for a narrative schema, in order to provide its normative element: we would not call something a narrative unless there is an interpreter involved.

This is not to suggest that the *individual elements* of the narrative would not *exist* without an interpreter. On the contrary, NRS2 - which captures part of the actual history of the subject - consists of *actual* causal relations among the relevant construed events. The point rather is that the narrative itself does not figure those relations and so does not exist without the interpreter. This makes sense - a narrative is a formal structure and there are many ways of describing a particular set of events or states, depending on our interest. So, for example, recall that the bad guys in *Martin Chuzzlewit* construed Tom's love for Pecksniff as a natural element in an idiot-type narrative; while we, the good guys, went with Dickens and generosity. Fortunately, in literature, we have a narrator to guide us in this way. In real life, I am afraid, Tom's narrative will have to be an interplay between Tom, the good guys, and the bad guys

and the feminist guys and the Marxist guys, and so on. This has the paradoxical result that the narrator of personal narratives that we have been hunting for, is after all the interpreter. Moreover, there is no single narrative for any individual or emotion (though each has a causal and reason-history)⁷⁹.

The conclusion may be a little disturbing, considering that the relationship between the actual narrative structure of a person's emotion and paradigmatic narratives looms so large in my notion of appropriateness. It may be suspected, in other words, that I am lapsing into deep subjectivism, which would be a *reductio* of my thesis. But I did not say that anyone's narration of a person's narrative would be accurate or would do for the appropriateness test. On the contrary, I repudiated the narrative which the bad guys imposed on Tom, and argued that we should adopt instead Dickens's envisaged narrative. It would be the same in real life. The more informed about the narrated a narrator is, the more we should heed his narrative in estimations of appropriateness. And of course, the question arises what counts as more informed and accurate, seeing that I deny that we have unmediated access to the world. But questions like these are not endemic solely to my thesis - they would also be a problem for anyone who casts appropriateness in terms of a one-to-one map between a construal and reality - so they cannot count as a *reductio* here. I am just more honest in following through the implications of a mediated experience: if one is interested in divine appropriateness, then one must (though one cannot) use NRS - a divine history - in evaluations of others' emotions; if one wants to talk about appropriateness accessible to humans, then one must use narratives - subjective, fallible, fallen things.

Conclusion

An emotion, then, is a mental state characterised by RS*, where the object-aspects of the Schema are cast in terms of concern-based construals, which are in turn embedded in a person's narrative; and the attitudinal aspect is cast in terms of phenomenology, expressive behaviour, and seriousness towards the construal.

An emotion is *intelligible* if the concern-based construal which is its object is also its cause; and as far as the agent is concerned the construal instantiates the formal object of the emotion.

⁷⁹ I am grateful to Paul Trompeter for showing me that, despite my avowed intentions, I was actually coveting a conception of narratives which figures *The Narrative of a person* and *The Narrator of a narrative*.

An emotion is *appropriate* if it is intelligible; and if its narrative *as an episode in one's overall narrative* instantiates (one of) the emotion's paradigmatic narrative structure(s).

CHAPTER 3

Introduction

The present chapter concludes the argument of this dissertation. In Chapter 1 I focused on predictions of a mental state on the basis of knowledge of the subject's other mental states. I argued that Simulation is less heuristically parsimonious and that Theory is necessary for predictions. This chapter integrates these insights with the analysis of emotions offered in Chapter 2. Part I shows that Theory is necessary for all emotion-*predictions*⁸⁰, and necessary and sufficient for predictions involving emotions which the ascriber has never experienced. Part II shows that Theory is necessary and sufficient for the *attribution* and *prediction* of all emotions. In the course of the chapter Theory emerges as more hospitable to the 'from the inside idea' than Simulation.

Part I: Predictions of emotions

Introduction

Two considerations showed Theory to be necessary for predictions in Chapter 1: (a) the simulator must feed as input into his own mechanisms for practical and theoretical reasoning those of the subject's mental states that are *causally relevant* for the predicted state; and (b) he must ensure that his input states are *relevantly similar* to the subject's states antecedent to the state predicted. Alvin Goldman has argued that objections from (a) relevance and (b) relevant similarity misconstrue simulation as theory-driven (and so beg the question), whereas it is in fact process-driven. This is how he explicates the notion of process-driven simulation:

if one person simulates a sequence of mental states of another, they will wind up in the same (or isomorphic) final states as long as (A) they began in the same (or isomorphic) initial states, and (B) both sequences are driven by the same cognitive process or routine (1995a: 85).

It should be immediately obvious that this response will not propitiate the Theorist. Part (A) of Goldman's formula is a reiteration of my worries, rather than their solution. The arguments for Theory's necessity in Chapter 1 were based precisely on considerations concerning (A). The

⁸⁰ Once I had written this chapter, I stumbled on Goldie's (2002) argument for the claim that Theory is necessary for emotion predictions. As far as I can see, we differ in the following ways. First, Goldie never spells out in any detail what his envisaged Theory is (which I did in Chapter 1 and will complete here). Second, he does not claim that Theory is *sufficient* for predictions, as I will in Part III below. Finally, he does not touch on attributions, and so my claim is not only greater in scope, but also more fundamental: attributions are conceptually and causally necessary for predictions (even on the Simulation model, as I argued, the simulator must ascribe input states before he embarks on the simulation). Finally, my argument is predicated on attributions' necessity for predictions, while his is not.

conclusion stands, I left (nearly) unquestioned part (B). In section 1 I dwell on part (A) *à propos* emotion-predictions. In section 2 I raise some doubts about (B) in the same context.

Section 1: Starting in the same states

What are we doing when Iago reveals his intention to “abuse Othello’s ear” and we predict that Othello will be jealous? Simulationists claim that we are “exploit[ing] [our] own motivational and emotional resources” (Gordon, 1996: 11). So presumably, when Iago announces his intention, we ‘transform’ ourselves into Othello, or imagine that we are like him in character and situation. We then imagine being confronted with evidence of infidelity, gathered by a trustworthy man, and - using our theoretical-reasoning mechanism - we predict that Othello will construe Desdemona as unfaithful. We then feed this construal, together with love for her, in our emotion mechanism, and predict that Othello will be jealous, based on our ‘pretend’ output.

Two aspects of the envisaged procedure invite attention. First, we are using two different mechanisms (first, theoretical, then emotional), and it is uncertain what, from Simulation’s armoury, can account for the transition between the two. I appreciate that the Simulationist need not be ruffled by this point - the transition, presumably will happen automatically, just as it would be if the simulator’s mechanisms were doing their usual job.

Second, and more damaging, the *input* we feed into our emotion-mechanism must be what Rorty called the *immediate* cause of the emotion (construal of Desdemona as unfaithful). But as shown in Chapter 2 (Part II, section 3), the causal efficacy of the immediate cause derives from the significant cause. I argued that the concerns involved in the construal, as well as its seriousness, constitute the significant cause. So I said that it is Othello’s construal *and* his love which cause the jealousy. This point is crucial for the envisaged simulation procedure: in order to come up with a prediction of *Othello’s* emotion, we must take into account a large portion of its causal history, both significant and immediate. Without feeding the mechanism with the *full* input, it would not function in a way which qualifies as a prediction of *Othello’s* emotion. Consider someone, who does not love his wife, learning that she is unfaithful. Two outcomes are possible: if he is not possessive, he would not be jealous; if he is, he would (as Iago is jealous, though in a very different way, cf. Chapter 1, Part III). Conversely, Othello could love Desdemona and believe she is unfaithful, and still not be jealous. Of course, these considerations suggest that love is neither necessary nor sufficient (even in combination with belief in unfaithfulness) for jealousy⁸¹. But invoking love is important for showing that we need

⁸¹ I am indebted to Peter Goldie for persuading me (in personal correspondence) that love is neither necessary nor sufficient for jealousy.

to consider Othello's broader characteristics in order to predict that *he* would be jealous, since it is these characteristics which play the principal causal role in his jealousy. But such considerations can only come from Theory. And Theory does not face the envisaged problem⁸², because it does not posit the use of a *mechanism*, and so the question of correct input (and so of the emotion's causal history) does not arise (more of this in Part II).

Now, I said that Iago's jealousy is different from Othello's, which may suggest that I am making the trivial point that in order to predict emotional *nuances* we need to know the individual history of a person. But my point is more fundamental than that: knowing the causal history is necessary for predicting that a particular person is in the grip of an emotion *type* quite generally, and only on the assumption that we are predicting by simulating. There are two reasons for this. The first is to do with the plethora of possible mechanisms by which one can arrive at the same emotion-type. I will develop this point in the next section which deals with mechanisms.

The second reason that the causal history of an emotion is necessary for a simulated prediction stems naturally from taking seriously the 'from the inside idea' and Simulation's insistence on *mechanisms* doing the causal work in predictions. As soon as we use mechanisms as the vehicle of prediction, the *causal* story of the predicted emotion becomes the only thing that matters. This is because a mechanism, by definition, is fed not with reasons but with *causes*. But now consider Heal's claim that psychological explanations and attributions are essentially first-personal and *reason-providing* (p. 5 above). As soon as this is granted and it is conceded that predictions through mechanism involve causes only, we must make space for reasons in a prediction. Since the mechanisms themselves offer no such space, *understanding* of the person is necessary either at a stage prior to the envisaged simulation or after it. But understanding after the simulation will not do, since we need to feed much of the causal story relevant for understanding into the mechanism. Understanding of the person's character and of the mental states antecedent to that predicted is then necessary *before* the simulation begins. Only such understanding, I submit, makes the input we feed into the mechanism reasons-laden.

It may be objected that I have neglected that the mechanisms in question are not just brute mechanisms like homeostasis mechanisms, but mechanisms for *reasoning* and emoting, and so are essentially fueled by reasons. I agree that reasons are essential, but the objection misses the point. When I think about *my* hypothetical thoughts, actions and emotions, the mechanism works with reasons naturally. This is because causes in my case *motivate* and so

⁸² I am indebted to Marius Vermaak for pointing out that the objection from multiplicity of possible causal histories of the same emotion could also be a problem for Theory.

*rationalise*⁸³ my thoughts, actions, and emotions. So, if I was Othello, being a jealous sort of person, my jealousy would not merely be *caused* by love and construal of the beloved as unfaithful, but also *rationalised* by the emotion and construal. But if I am not a jealous person, my beloved's unfaithfulness can neither motivate nor rationalise my jealousy. And so, when I come to simulate Othello, I need to *understand* his character *and* know the causal story of the emotion, in order for my ('I' = Othello) jealousy to be rationalised. But such understanding cannot come from further simulative use of those mechanisms, because all the problems with relevant similarity developed in Chapter 1 resurface, and hence the simulationist would face an infinite regress. The upshot is that Theory is necessary for predictions at least at the input-feeding stage of the envisaged simulation.

Section 2: Using the same mechanism - unexperienced⁸⁴ emotions

I argued in Chapter 1 (Part II, section 2) that Heal can conclude that the simulator and the subject would arrive at the same solution to the quantum problem, only by assuming (a) that they start in the same initial state; and (b) that they share specific (quantum) applications of the general theoretical-reasoning mechanism. Section 1 above dealt with (a), which is also part (A) of Goldman's formula, in the context of emotion-predictions. It is time to turn to the second assumption (Goldman's (B) condition).

I have already embarked on this journey in section 1. Recall that there I anticipated being charged with the trivial claim that an emotion's causal history is necessary for predicting emotion-*nuances*. I said that there were two steps to arguing that I am innocent of triviality. First, I showed that understanding is necessary at the *input* stage of the simulation procedure. Second, I promised to parry the charge by exposing Simulation's problems with *mechanisms*. The charge of triviality can also be cast in terms of mechanism. The argument (for the trivial conclusion) that my opponent could offer on my behalf would be this⁸⁵: (i) because there are many kinds of jealousy, (ii) there are many mechanisms for jealousy. (iii) We thus cannot assume that the mechanism we are actually using in simulation is the right one, and so (iv) we need understanding of character in order to use the right mechanism. But (v) such understanding requires Theory.

I agree with (ii) to (v), but the move from (i) to (ii) needs serious modification. It is this move which would make my point trivial, because all that follows from (i) to (v) is that we

⁸³ I used to think that causes *justify* actions, thoughts, and emotions. I am indebted to Eusebius McKaiser for urging that justification is too strong.

⁸⁴ 'Unexperienced emotions' will be my (I appreciate graceless) shorthand for 'emotions which the ascriber has never experienced'.

⁸⁵ I am indebted to Marius Vermaak for showing me that my argument can be thus misconstrued.

need Theory to predict emotion-*nuances*, and of course no one would disagree with this. It is not, I submit, because there are *many kinds of jealousy* that there are many mechanisms for getting into jealousy. Rather, there are many mechanisms for arriving at the general type of emotion jealousy, because of the analysis of emotion offered in Chapter 2. First, a person who only *sees* (in the serious-construal sense, Part II, section 3) his wife as unfaithful can be just as jealous (both qualitatively and quantitatively) as the person who knows his wife to be unfaithful. Second, a person who does not love his wife but sees her as property, and is possessive about his property, can be just as jealous as the person who loves his wife. The first consideration shows that *doxastic commitment* of different degrees can figure in the causal story of the *same emotion*. The second shows that different *concerns* can figure in the causal story of *the same emotion*. It is obvious that the vast number of possible permutations of concerns and construals lead to a plethora of very different mechanisms for getting into the same emotion - jealousy. Now steps (ii) to (v) follow, and compel the conclusion that Theory is necessary for predicting that a person is in the grip of a *general type* of emotion, like jealousy.

Consideration of unexperienced emotions will render Simulation's assumption of mechanism similarity even more dubious. My argument will draw on such emotions in two contexts⁸⁶: (1) any mental-state prediction which *involves* the *ascription* of unexperienced emotions; (2) predictions *of* unexperienced emotions.

1. *Mental-state predictions which involve ascriptions of unexperienced emotions*

A young woman is presumably incapable of paternal love⁸⁷. What does she do when she is watching *King Lear*, and predicts that Lear's heart will be broken by Gonerill's and Regan's cruelty, and grateful for Cordelia's forgiveness? Simulation must say that she is using her own emotional resources. There are four main aspects to the input she must feed into her

⁸⁶ The distinction mirrors Gordon's distinction between concepts involved in the process of prediction and the concept of the predicted mental state (Chapter 1, Part I, section 1).

⁸⁷ Marius Vermaak has suggested that we are capable of all emotions, and so it is inaccurate to say that the young woman is *incapable* of such love. Now, all that counts for my argument is that the ascriber has *never* experienced the emotion in question, or is somehow precluded from experiencing it. But it is worth engaging with Vermaak's comment for the sake of completeness. The quick response is that, by definition, a *woman* cannot experience *paternal* love. But this is rather facile. A better response would be to claim that if she is capable of *experiencing* it, she must have the *capacity* for love in general *and* an *imagination*. The imagination itself will have to be aided by some Theory. This is all that counts for my argument. And notice that this line of thought is different from the arguments offered throughout the chapter, in that what is at stake is not the capacity for *predictions* and *attributions*, but the capacity for *experiencing* an emotion. I am adding this caveat for the benefit of anyone who might think that my requirement of a capacity for love when it comes to capacity for *experience*, conflicts with my claim that Theory requires no such capacity (as developed in the discussion below of Regan, Gonerill, and Edmond).

mechanism: Lear's concern with sustaining his royal dignity; his expectation that his daughters will observe their "offices of nature, bonds of childhood" (2.4.171); his entitlement to respect due to his age (2.4.183-4); and his paternal love. Suppose, contrary to the conclusions of Chapter 1, that our young female viewer can use her practical and theoretical reasoning mechanisms for the first three aspects, without recourse to Theory. What about the fourth? Since she has never experienced paternal love (in the sense that she has never paternally loved anyone), her emotion-mechanism can do nothing with it, because it does not 'recognise' the emotion and does not 'know' how it functions causally. She must then invoke some Theory in order to categorise it. The piece of theory would be of the sort: 'paternal love is a type of love, so I need to use my capacity for love in the envisaged procedure'. But not just any sort of love will do. Awareness of a certain sort of responsibility is intrinsic to paternal love, and it is not the same responsibility as the child's towards her parents, or a lover's, or a friend's. In order to feed her mechanism with paternal love, then, our young woman must *categorise* it as love which involves a particular sort of responsibility and protectiveness. Only once this is done can she use her capacity for love (which is the appropriate mechanism) to simulate a prediction of Lear's despair.

The discussion so far suggests that unexperienced emotions can still be simulated in order to arrive at a prediction of a mental state which causally involves them, though the simulation requires some theory. If we consider the Edmond-Gonerill-Regan trio, however, it will become obvious that simulation has no place in predictions involving the ascription of unexperienced emotions. What did the mechanism work in the simulation procedure envisaged above, was the young woman's *capacity* for love *simpliciter*. In *King Lear*, however, none of the members of the trio is capable of love, yet they can all competently attribute it and handle predictions involving it.

a. Incapacity for love

After professing their love for Lear at the beginning of the play (in order to get their portions of the kingdom), the two sisters never show so much as affection for him. They withdraw the last privileges which nourish his dignity. They shut their doors to him, abandoning him "To wage against the enmity o' th' air" (2.4.202) during a storm that even Odysseus was rarely expected to brave. Their capacity for sibling love is shown in Gonerill's poisoning Regan and Regan's stabbing Gonerill to death. Their capacity for 'romantic' love is exhausted in a race for Edmond, in which it is unclear that attainment of him - rather than triumphing in sisterly rivalry - is the purpose. Edmond, on his part, appears just as gloriously equipped for love. He betrays his brother to exile; does everything to ensure his father's blinding (though his

tenderness shrinks at actually witnessing it); and love is never so much as a consideration in his cogitations about which of the sisters to espouse⁸⁸.

b. Competence in ascriptions and predictions

Despite their incapacity for love, all three are perfectly capable of *ascribing* not only love, but paternal love. Gonerill and Regan know that Lear loves Cordelia most, and, initially puzzled by his repudiation of her (1.1.281), soon explain it (1.1.280-290) in terms which attest not only to their ability to ascribe it correctly, but also to their knowledge of what the emotion involves. Edmond ascribes paternal love to his father Gloucester (1.2.17).

Moreover, the trio is capable of using paternal love in *predictions*. Edmond knows what beliefs to induce in Gloucester in order to famish his love for Edgar and strengthen it towards himself (1.2.19-21). Finally, Gonerill predicts that Lear will be indignant and bereft when she refuses to continue housing all of his followers. She knows also that Lear will seek Regan's support. Indeed, Gonerill is so good at predicting all this that she has written a letter to Regan, informing her of Lear's approach, *before* the scene takes place (1.4.285).

We have, then, as in the case of Iago (Chapter 1, Part III), three people who are perfectly incapable of love, and yet perfectly capable of attributing *paternal* love, and using it in predictions. It should be obvious now that simulation has no place in predictions which make use of unexperienced emotions, since the relevant emotional capacity or mechanism which allegedly drives the envisaged simulation in predictions is just not in place. The conclusion can be extended to predictions *of* unexperienced emotions.

2. Predictions of unexperienced emotions

In the previous sub-section, I showed that Theory is sufficient for predictions which *involve* unexperienced emotions. Now I want to show that Theory is sufficient for predictions *of* unexperienced emotions. For this we will need an example of a person who is incapable of simulating the causal history of a particular unexperienced emotion, and yet is capable of predicting that another will experience this emotion. This would show Theory sufficient for

⁸⁸ It may be argued that Edmond's repentance before he dies attests to his either having had the (dormant) capacity for love all along, or at least for having some future potential for it. Only the first contention would undermine my argument. The way to parry this possibility is to notice that the repentance is very abrupt and a little too close to the prospect of Edmond's facing his Maker, to be deep, or indicative of dormant love or goodness. Compare his repentance with Leartes's at the end of *Hamlet*. Leartes's appears to the audience a genuine conversion, partly because we never get a sense of thorough evil in him. There is no doubt that Edmond, by contrast, is constructed as the paradigm of evil, which makes an inchoate capacity for love unlikely. I am indebted to Wendy Jacobson for a discussion on this topic.

such predictions. I argued in the last sub-section that the *capacity for love* would be necessary for a simulation-*attribution* of paternal love. But since predictions are conceptually and causally predicated on attributions, it follows that for a simulation-*prediction* of paternal love, the capacity for love is necessary. Now consider Regan, Gonerill, and Edmond again. I argued that they are incapable of love. Seeing that they are capable of all sorts of predictions *involving* the *attribution* of paternal love, there is no reason to suppose that they cannot predict that a loving person, to whom a child is born, will feel paternal love. It follows that people without the capacity/ mechanisms which would do the work in a simulation-prediction, are capable of predicting the emotion nonetheless. It further follows that Theory is sufficient for predictions of unexperienced emotions.

It may be objected that what did the work in this argument is the incapacity for certain emotions *involved in the causal story* of paternal love, rather than the unexperienced emotion *itself*. And so, the conclusion holds only for predictions of unexperienced emotions whose causal mechanisms also involve a *further* unexperienced emotion. The objection is misguided. According to the holism of the mental that I endorsed in Chapter 1, a mental state (and so an emotion) is not an item detached from other mental states. The insight applies to mental states as (a) concepts and (b) experiences. Mental state holism, then, implies that (a) one could not have the concept of an unexperienced emotion without other mental-state concepts, amongst which would be emotion-concepts; (b) one could not experience an emotion without being capable of either experiencing other related emotions or at least of having the concepts of these emotions.

Conditions (a) and (b) imply, in turn, that the causal history of an unexperienced emotion will typically figure other emotions which the ascriber has not experienced. So the argument developed in this subsection stands:

- (1) The causal history of a particular unexperienced emotion *typically* involves other unexperienced emotions;
 - (2) If a person (Edmond, Gonerill, Regan):
 - a) is incapable of an emotion (love) required for the simulation-prediction of an unexperienced emotion (paternal love),
 - b) but can handle predictions of it,
 - (3) That person cannot be using Simulation.
 - (4) But the choice between Theory and Simulation is exhaustive.
- Therefore, Theory is sufficient for predictions *of* unexperienced emotions.

Conclusion

The last two sections show, first, that Theory is necessary for predictions of emotions. Second, Theory emerges as necessary and sufficient both for predictions of unexperienced emotions and for predictions which involve the attribution of unexperienced emotions. The second insight suggests a sketch for an argument for Theory's sufficiency for predictions of *all* emotions. The argument would have to show that there is no difference between predicting emotions that one has experienced and emotions that one has not. I will develop this point once I have discussed attributions. The reason for my present shyness is that at this stage the only way to establish that there is no such difference would be by an appeal to phenomenology. Now, since there does seem to be a phenomenological difference (there seems to be much more theoretical thinking in the case of unexperienced emotions) one would have to argue that the difference is one of degree rather than of kind. But now recall that the Theory we are supposed to be using, and our actual use of it, can be partially tacit or inaccessible to consciousness. In light of this, appeals to phenomenology will not be strong enough to settle whether a difference is one of degree or of kind. So an argument for Theory's sufficiency needs to be based on conceptual considerations. This is how I presently argue in the context of emotion-*attributions*. I then revisit predictions in section 3 below.

Part II: Attributions of emotions

Introduction

In this part I extend to emotion-*attribution* the conclusions about prediction attained so far. I argue that Theory is necessary (section 1) and sufficient (section 2) for such attributions. In section 3 I argue that it is also sufficient for predictions. The arguments for necessity in section 2 develop in response to a stock defence Simulationists deploy against objections from relevance and relevant similarity. The arguments for sufficiency are founded on the Relational Schema developed in chapter 2, and on the insight that having the concept of an emotion as captured by RS* is sufficient for competent attributions. Finally, Theory's sufficiency for predictions follows from the above conclusions.

Section 1: Theory's necessity for emotion-attributions

First, it should be obvious that the objection from mechanism developed in part I, section 2, applies automatically to attributions. As already shown, Edmond, Regan, and Gonerill are perfectly capable of ascribing paternal love despite their incapacity for *any* love. This means that they cannot be using their own emotional resources in an attribution and so cannot be simulating their fathers. What of the claim, developed in section 1, that the ascriber must

replicate the mental states causally relevant to the emotion predicted, and so must import Theory at least at the beginning of the simulation procedure?

Heal considers this objection in the context of attributions and argues that such an objection

misdescribes the direction of gaze of the [simulator]. He is not looking at the subject to be understood but at the world around that subject. It is what the world makes the [simulator] think which is the basis for the beliefs he attributes to the subject (1995a: 48).

In this section I follow Heal's gaze towards the objects of *emotions*. I argue that turning our gaze to their world-aspect exclusively (I dub this 'Sunflower Simulation') would preclude us from ever non-arbitrarily attributing emotions to another.

Simulation, I submit, faces the following dilemma:

Either (a) its emotion-ascriptions take into account the entire objects of emotions, in which case the ascriber's gaze is not directed solely at the world (since a construal is irreducibly personal), and Theory needs to be invoked;

or (b) the gaze is directed solely at the world and the actual object of emotion is disengaged from the ascription, in which case it is unclear that what Simulation is doing is emotion-ascription to *another*.

The last quotation shows that Heal is, understandably, reluctant to countenance the first horn. Similar misgivings are expressed by Robert Gordon. So I turn to the second horn. Gordon offers an illuminating example explicating the notion of turning our gaze exclusively to the world in an ascription (his term is "total projection"). He invites us to imagine that we are walking in a mountain with a friend. Suddenly the friend recoils back down the path. How do we explain his behaviour? We ascribe fear to him. But how?

You follow him, looking over your shoulder *to search the environment* for an explanation...You look for salient features in the middle distance, particularly for *menacing, frightening* things...The question for you is something like this: 'What is it about these environing rocks, trees, animals, and so forth, that would explain his suddenly turning back?' And the question is understood to presuppose that your friend is *aware* of these very objects, and that whatever it is *about* these objects that constitutes the *explanans* you are seeking is something *known* to him (1995b: 102-103, italics in original, my emphasis).

Now, for the Sunflower view to be a solution to the problem of relevant mental states, which motivated it, *and* to differ non-trivially from Theory, it must claim not only that the underlined question is *necessary* for ascription but that it is also *sufficient*. This is where Simulation faces the second horn of the dilemma. To see why, consider the analysis of the object of emotions developed in chapter 2. It was captured in RS*:

$R(S t f a c p m s)$,

where R is the emotion type, S is the subject, t the target, f the focal property, a the motivating aspect, c the cause, p the proposition specifying the ground of the emotion, m the aim, s the characteristic feeling towards the object of the emotion (including seriousness towards the construal and characteristic phenomenology).

Five of the parameters characterise the emotion's object. We need to take all five seriously if we are to meet Heal's requirement of substantive content (Chapter 1, Part II, section 1). But it is immediately obvious that out of the five only the target can be accommodated by the Sunflower view. The cause either has an irreducibly personal component (if I am right that it is typically the construed object) or is non-rational (something like too much coffee). The focal property, while it can (and usually is) an actual property of the target has an irreducibly personal element - it is focal *for* the subject to whom we are ascribing the emotion. The motivating aspect, considering that it is a function of cause and focus has an equally personal component. Similarly with the propositional object, which captures the reason for the emotion. Significantly, it is the focal property, the cause, and the motivating aspect, that *explain* any particular emotion and thus justify our ascription of this emotion. I return to this point presently.

I have not yet mentioned what scared Gordon's friend. It was a grizzly bear. Now, obviously when the object of fear is so objectively dangerous, we need no fine-grained analysis of its individual aspects. Indeed, we can just turn our gaze to the world and discern cause and object all in one. So it is not surprising that Gordon can perform a Sunflower-simulation and ascribe the emotion correctly. But now suppose that next to the grizzly bear there is a burnt tree, dead in a pool of burnt grass. And suppose, too, that Gordon's friend finds grizzly bears adorable and not frightening in the least. Burnt trees, on the other hand, are the ultimate source of horror to him, since when he was a child a tree in his garden burnt down and when the fire started he was sleeping in his tree-house in that very tree. And suppose also that Gordon knows his friend's attitude to both grizzlies and burnt trees. Now let us replay the hike. Gordon's friend recoils back down the path. Gordon turns around to look for the cause in the environment. He sees grizzly, he sees burnt tree. What emotion *must* he Sunflower-ascribe to his friend? Fear of the grizzly bear. Why? Because on the Sunflower model he has only recourse to the environment and grizzlies are objectively dangerous, while burnt trees are not objectively dangerous or horrifying. But, by assumption, Gordon knows his friend's aversion to burnt trees and fondness for grizzlies, and so what he would *actually* ascribe to him is horror of the burnt tree.

This suggestion has several interesting consequences. Firstly, it should be noted that I am *not* arguing that Simulation's problem consists in the possibility of a particular simulated ascription's being wrong⁸⁹. This would be gratuitous and pointless⁹⁰: we get other-ascriptions wrong all the time. My point rather is that whether the simulator gets the ascription right or wrong is an *arbitrary* matter since he cannot know - without recourse to the mental states of the subject - *whether* and *why* he's got it right or wrong. And I argued that understanding is necessary for emotion-predictions above, since the mechanisms envisaged by Simulation are causal ones, and so understanding prior to the simulation procedure is the only thing that allows room for reasons. The point applies to attributions. Understanding, in turn, requires recourse to the mental states of the other. But, of course, adverting to the mental states of the other reconjures the first horn of the dilemma, with which I opened this section, and hence all the problems from relevant similarity discussed previously.

In order to develop my argument for this claim, let me consider a possible response to it. It may be objected that my point does not essentially depend on any personal aspect of the emotion's object, and still less on the fine-grained analysis of emotional objects offered in Chapter 2. The reason Gordon got the ascription wrong, after all, was that he got wrong the *target* of the emotion (the tree out there instead of the grizzly out there), rather than any specific properties of it. And recall that I agreed that the target is one of those things about the emotion for which the Sunflower view *can* account.

The objection is too quick for three reasons: Gordon's ascription method (1) confuses cause and object; (2) is either arbitrary or employs illegitimate standards of appropriateness; (3) could never be an ascription of an emotion which is intelligible but inappropriate.

1. *Confusion of cause and object*

My analysis distinguishes the *cause* from the *object* of the emotion, which the Sunflower view cannot accommodate. In Friend's case (and in typical cases) it so happens that they coincide (as they must for Gordon's story to work). But it may have been that what *caused* Friend's fear was a general timorous propensity enhanced by an overdose of coffee. In this case, when turning his gaze to the environment and seeing the grizzly, Gordon would attribute correctly

⁸⁹ I am grateful to Marius Vermaak for pointing out that my argument may be thus misconstrued.

⁹⁰ Though the point is developed non-trivially by Nichols *et al.* They argue plausibly that the fact that we make mistakes in attributions favours Theory. The argument turns on the consideration that Simulation relies on *shared* cognitive mechanisms, but if these mechanisms are shared, then we cannot explain attribution errors (except by discrepancies in input, which Nichols *et al.* eliminate in several experiments, and still subjects get attributions wrong). Theory, by contrast, is perfectly equipped for errors: if the ascriber's theory includes false information, or he is not using the relevant bits of information, then it makes sense that he gets the ascription wrong. (Nichols *et al.*, 1996: 50-52).

fear but misattribute its cause and hence motivation (though he gets the *putative* object right). Gordon may respond that as long as he gets the emotion right, the cause does not matter. The response will not do for three reasons. First, it would rather betray the purpose of other-attributions – to determine what the other is thinking/ feeling about the environment. Second, as I argued in Part I, Simulation must feed causes into the relevant mechanisms, and not objects. Third, the response would be dissonant with Gordon's overall project, which is to show that Simulation is more propitious than Theory for *explaining* mental states (1995b: 112). In fact, his accusation of (a rather straw-man version of) Theory is precisely that Theory may get the ascription right without knowing why.

2. *Arbitrariness or illegitimate deployment of standards of appropriateness*

The second reason that the appeal to the target in the environment is too quick is that the target of Friend's horror was just a tree. The focal property (and presumably the motivating aspect) was its being burnt. There are two reasons for Gordon's simulation not ending up with a horror-attribution:

- a. He was, as he admits, looking for "menacing, frightening things" rather than horrifying things. This is a telling point in itself. Gordon seems to have used a theory about what causes people to recoil on a mountain path (fear, rather than pleasure or horror).
- b. The second reason he did not attribute horror to Friend when he was examining the environment, is that for him being burnt is not a *focal property* of a tree, and certainly not a property which can motivate horror. It is because of this that he gets both object and emotion wrong: he has assumed too much in the simulation: when he turns his gaze to the environment he is already looking for an *appropriate object* of (what he has decided is) the *appropriate emotion* explaining Friend's behaviour. The appropriateness of the object can be either objective or what *he* (but not what his friend) would deem appropriate. If it is objective, he is using theory. If it is appropriate for him, he is either using theory to set criteria of relevant similarity between himself and Friend, or he is not attributing anything to Friend.

3. *Inability to ascribe intelligible but inappropriate emotions*

This point suggests the third reason the target-objection is too quick: Sunflower-Simulation can never attribute an emotion which is *intelligible* but, nonetheless, *inappropriate*. Friend's horror of burnt trees is a borderline case. His fondness for grizzlies would make for a better example. Suppose Gordon saw the grizzly first, and was trying to predict his friend's emotion and its expression. He could not, without recourse to his knowledge of Friend's fondness, predict the

enthusiasm, the tear in the eye, the attempt to stroke Grizzly. It is clear, however, that if emotion-attributions are to explain the behaviour of *someone in particular*, rather than explain *appropriate* behaviour or the possible behaviour of the *ascriber*, a model of ascriptions must canvass intelligible but inappropriate emotions.

This point is not meant merely to expose Simulation's (very) limited scope of ascriptions. Since appropriateness itself is founded on intelligibility, if a model of ascriptions cannot make sense of the distinction, then it cannot make sense of intelligibility, and so the notion of appropriateness itself becomes arbitrary. This has two further consequences. First, we could never judge an emotion utterly irrational or self-deceived, and still less talk of our responsibility for emotions. Second, if we cannot make sense of intelligible but inappropriate emotions, then we have effectively countermanded Heal's requirement of substantive content.

4. *Objections*

Now, Gordon admits that total projection is the default mode, and we do adjust for relevant differences when it seems necessary. Once this adjustment is taken into account, it may be argued, Simulation can face all three problems laboured above. But this response courts an infinite regress. The discussed problems are at the heart of Simulation. Recall that the Sunflower view was an attempt to handle objections from relevance that Simulation faced. It will be of no use invoking general Simulation now in order to fix the problem with Sunflower Simulation.

It may be further objected that I am simply begging the question against Sunflower Simulation⁹¹: my analysis of emotions is a *Theory* account of emotion, the envisaged objection goes, and it is no surprise that it would show up Simulation. I resist the imputation on the following grounds. First, although de Sousa's account (on which mine is based) historically coincides with a Theory perspective, I have not just assumed his position but argued for, and refined, it. Moreover, the arguments were motivated by problems which had nothing to do with the Theory-Simulation debate. If the arguments worked, my version of de Sousa's position, is substantiated independently of Simulation-Theory considerations. Second, the arguments against Sunflower Simulation should be viewed in the context of the general arguments against Simulation developed in Chapter 1. These arguments raised perfectly general doubts about Simulation, independent of any Theory-conception of mental states. Finally, a Theory/Simulation vision of the nature of mental states is neither necessary nor sufficient for a Theory/Simulation answer to the question of predictions and attributions. So a Theory-conception of emotions does not entail a Theory view of attributions and predictions. This point is granted

⁹¹ I am indebted to Marius Vermaak for this objection.

not only by light-weight champions of Simulation (like Heal), but by Gordon himself, who thinks that Simulation is the correct answer to all three levels of the Basis Question (nature of mental states; concept mastery; and explanation, prediction, and attribution).

Sunflower-Simulation, then, cannot offer a satisfactory account of how we attribute emotions to others. In fact, it is not clear that the view is concerned with attributions of *actual* emotions to others, rather than with statements of what others *ought* to feel if they were the ascriber. We are then still faced with the question how Simulation proposes to conduct attributions and predictions which do justice to the point of view of the agent, but which do not advert to Theory. In the absence of an answer, Theory proffers a promise of greater heuristic parsimony and hospitality to the from-the-inside idea.

Section 2: Theory's sufficiency for emotion attributions⁹²

In Chapter 1 I mentioned that Heal distinguishes three loci of the Simulation-Theory debate:

- (a) *understanding* of what is involved in judgements about others' thoughts in general;
- (b) *attribution* and *prediction* of a particular thought on the basis of knowledge of the subject's *behaviour* and *circumstances*;
- and (c) *attribution* and *prediction* of a particular thought on the basis of the ascriber's prior knowledge of the subject's *other mental states*.

I also mentioned that she argues that Simulation is the apt approach to (c) and may play some role in (a) and (b). So far in this chapter I have argued that Theory is *necessary* for all emotion-attributions and predictions, because they require understanding of the emotion, which in turn requires Theory. In the present section I show that such understanding, and hence Theory, is *sufficient* for all emotion-attributions. The argument unfolds as follows: First, I show that knowing the revised Relational Schema for a particular emotion is sufficient for understanding what is involved in judgements about others' emotions. Second, I argue that such understanding plus knowledge of paradigmatic narratives is sufficient for attributions. It follows that Theory (represented by RS* and paradigmatic narratives) is sufficient for attributions.

1. *Understanding what is involved in judgements about another's emotion*

In Chapter 2 I argued that emotions are mental states characterised by RS*, where the object-aspects of the Schema are cast in terms of concern-based construals, which are in turn

⁹² I am indebted to Marius Vermaak for showing me that my accounts of emotions and of the Theory-Simulation debate could be extended to Theory's sufficiency for attributions.

embedded in a person's narrative; and the attitudinal aspect is cast in terms of phenomenology, expressive behaviour, and seriousness towards the construal.

What do we need to know in order to understand what is involved in a judgement about another's emotion? The most obvious candidate is having the concept of the emotion. A particular version of RS* captures the concept of a particular emotion. So, for example, the RS* for envy⁹³ tells us that envy involves (a) two targets - the envier (A) and the envied (B); (b) a construal of B as having a desirable focal property (X), and a construal of A's either lacking X or having an inferior X; (c) A's concerns of the sort, 'having X is vital for my self-conception/ well-being' (if the envy is non-malicious) or of the sort 'B's *not* having X is vital for me' (if the envy is malicious⁹⁴); and (d) expressive behaviour, like going green when B flaunts his X, schadenfreude-smile when B breaks his X. Presumably, we do not need to know what the characteristic phenomenology of an emotion is, or how serious the seriousness involved is, in order to have the concept⁹⁵. As argued in Part I above, Gonerill, Regan, and Edmond are incapable of love and so lack knowledge of its phenomenology, yet have the concept, since they attribute it correctly (concept-possession being necessary for attributions⁹⁶). And as I argued in Chapter 1, Iago does not appreciate how serious the seriousness involved in jealousy is, yet can attribute it competently (Part III, section 2 above) and so has the concept.

What, over and above having the concept, do we require for understanding what is involved in attributions? I contend nothing. To see why, consider the two kinds of attributions Heal distinguished.

2. *Attributions on the basis of behaviour and circumstances*

Recall that the RS* features the aim of the emotion. The aim, in turn, was defined in terms of immediate expressive behaviour. I distinguished, after Velleman, two aspects of this behaviour: 'mere happenings' and 'mere activities' (Chapter 2, Part III, section 1). I also said that every emotion has a formal aim, constrained by the emotion's biological function and origins. The constraint would be absolute on *mere happenings*: our heart cannot palpitate with depression and our facial muscles cannot contract in perfect cheerfulness do what we may, since we do not have the requisite control. There are certain formal constraints on *mere activities* too, though it is less clear that they are grounded in evolutionary explanation: we do not jump from

⁹³ This sketch of envy is based, very crudely, on Goldie's elaborate analysis (2000: 221-224).

⁹⁴ I owe the distinction between malicious and non-malicious envy to Goldie (2000: 26).

⁹⁵ Again, I am not arguing that full mastery of the concept is necessary (cf. fn. 31, p. 35).

⁹⁶ This point emerges from Heal's two caveats on substantive content, and is conceded even by Gordon. Recall that he granted that the concept of the mental state predicted was necessary for 'genuine, comprehending attributions' (Chapter 1, Part II, section 1). The evil trio engage in precisely such attributions (p. 83 above).

depression, and we do not slouch from joy. Knowing the correct relation between an emotion type and its expressive behaviour, as represented by RS* is sufficient for attributions. If your eyes dilate, your mouth opens and you cringe, I know that you fear something. Of course, you can (more or less successfully) dissimulate all this, but all that this observation shows is that we can mislead each other about our emotions. That we *can* so mislead each other, in turn, shows just how heavily our attributions rely on behaviour.

The third component of expressive behaviour - actions - will be a little more troublesome, though. I said that actions are not part of RS*, but rather of the narrative which embeds the concerns and construals involved in the emotion. But recall that we are not attributing on the basis of the agent's behaviour in isolation from his circumstances. So, for example, if we want to explain a man's murdering his wife by attributing an emotion to him, we can, without recourse to his circumstances, give a range of emotions compatible with his action: jealousy, anger for burning his pants when ironing them or for being pregnant, covetousness of her money, irritation at her high-pitched and ever-busy voice, and so on. Knowledge of his circumstances and character will no doubt illuminate the choice. If there was another guy in the bed in which the husband killed her, then it was jealousy; if he has strangled her with a pair of pants which have a large iron burn on them, then it is anger at her having burnt the pants; if there is shattered crystal around the house, then it was irritation with her high-pitched voice, etc. We would also need to know a bit about the husband's character: is he the sort of person who can get *that* angry, irritated, or jealous?

Notice, however, that his character and circumstances are not external to the conception of emotion developed in Chapter 2: they are embedded in the notion of narrative structure. Now, this is not to say that one needs to know the narrative of a person before one can attribute an emotion. As I argued in this chapter, this was only necessary if we are attributing by simulation (Part I, section 2). Rather, knowing the circumstances, and behaviour of a person is sufficient for attribution, because both circumstances and behaviour are intrinsic to the structure and concept of any particular emotion. Knowing the concept of an emotion, then, involves knowing some paradigmatic narratives figuring (a) paradigmatic character types, as well as (b) paradigmatic situations which elicit particular emotional responses.

It may be objected at this point that my account raises the problem I imputed to the Sunflower view (section 1 here), namely, that the model cannot accommodate attributions of *intelligible* but *inappropriate* emotions. Since I have said that the actual narrative of the emotion does not matter, the envisaged objection goes, I am still attributing on the basis of a paradigmatic narrative, and so can only ever attribute *appropriate* emotions. But the objection is misguided. In Simulation, knowledge of a particular narrative is necessary in order to do the

causal work in the ascriber's mechanism. By contrast, on the Theory model, we do not need the actual causal story, because we are not using a mechanism. Further, room for reasons is not made by knowledge and understanding of the other's causal history: reasons are in RS* from the beginning, under the aegis of motivating aspect. The function of paradigmatic narratives is to guide our knowledge of the predicted subject. So our theory tells us that a jealous or choleric sort of person is more likely than an irenic sort to murder his wife, upon discovering her in the arms of the rival. We then judge on the basis of that and knowledge of the subject, what emotion to attribute to someone in particular circumstances. We are attributing an emotion that would be appropriate for someone of that character, and *not* as the Sunflower Simulationist did, an emotion which is *objectively* appropriate, or appropriate *for the ascriber*.

Now I seem to have gotten into further trouble: I have invoked character and so have subverted my claim that behaviour and circumstances are sufficient for the attribution. But this is not quite the case. First, speech is part of a person's behaviour and we often glean much from a person by what he tells us about the world and himself, and by the way he tells it. Secondly, we learn that a man is choleric or jealous or whatever by observing his *behaviour*, in the first place. And we can garner enough from speech, mere happenings, mere activities, and actions, to have at least some idea of his character purely on the basis of this inclusive sense of behaviour. Knowledge of behaviour, circumstances and character (derived from further behaviour) together with a Theory which relates these elements through generalisations (paradigmatic narratives), is sufficient for an attribution. Obviously, the less we know the person, the more generalisations we will use, and the less reliable the attribution would be. But this is consonant with our experience of attributions.

Notice, however, that the account does not require that we have experienced the emotion in question, though whether we have will obviously make for better grasp on the emotion and so for more comprehensive attributions. Experience of the emotion was only necessary for simulating, because there we were using mechanisms and so needed to make sure that we have the relevant mechanisms. The sufficiency of knowledge of behaviour for attributions is evinced in our everyday practice. When we ask someone why he did something, we expect that he cite mental states, one of which is often an emotion. In fact if we asked our husband from above why he killed his wife, and he did *not* cite an emotion (let us say he merely said 'Because I believed she was unfaithful') we would be either utterly baffled or would infer his jealousy (whether we are capable of jealousy or not).

3. *Attributions on the basis of knowledge of the subject's mental states*

a. *Antecedent mental states*

Considering everything said so far, it will be hardly contentious that Theory is sufficient for attributions on the basis of knowledge of the other's mental states antecedent to the attributed state. Recall that the causal story of an emotion consists of other mental states (concerns and construals), and the causal story itself is partly constitutive of the RS*. Knowledge of the immediate antecedents, or even of *their* causal antecedents, together with generalisations relating particular concerns and construals to particular emotions, will guarantee capacity for attribution. The *causal* story, of course, is not sufficient for explaining the emotion. But RS* caters for reason-explanations by virtue of its featuring motivational aspects. Unfortunately, attribution on the basis of other mental states *antecedent* to the emotion is not an attribution but a prediction. I return to this point in section 3 below.

b. *Consequent mental states*

RS* does *not* seem to cater for mental states consequent to the emotion attributed (though it does cater for consequent *behaviour*). So, knowledge of these states, together with RS* would not seem sufficient for attributions. We do, however attribute on this basis. Knowing that a woman's (call her Fidelia) belief in her husband's faithfulness persists in the face of evidence to the contrary, we are quick to pronounce that she loves him too much or that she seeks emotional security.

One way of maintaining Theory's sufficiency in the face of this apparent inadequacy of RS*, is to appeal to the action component of behaviour. If we have granted that we can adequately attribute emotions on the basis of action, and notice that action has an irreducibly mental component (intention), we have in effect conceded that when attributing on the basis of action, we are also attributing on the basis of a mental state. This, regrettably, will not do. Knowledge of action is *sufficient*, and not necessary for knowledge of the mental state. It is also sufficient (together with other relevant knowledge) for attributions. So to conclude, on the basis of that, that knowledge of a mental state is sufficient for attribution would be a simple logical error.

A more promising strategy for sustaining Theory's sufficiency in the context of attributions on the basis of knowledge of consequent mental states, is to appeal once again to narratives. Now, it is true that narratives have a historical, rather than futuristic role. They offer backward-looking explanations of emotions: a person's narrative explains (and so justifies the attribution of) a particular emotion. But now consider that a person's narrative encompasses his mental states other than emotions *and* tells us of the person's character. In

light of this, we can use the patterns that emerge from the narrative (patterns correlating previous emotions and other consequent mental states), in order to attribute an emotion on the basis of knowledge of its consequent mental states. So suppose that Fidelia from above explains away what is a sure sign that her son has committed a crime. How do we explain her belief that he is innocent? Well, one way is to call her stupid or utterly irrational. Normally, however, we are more charitable (albeit ambiguously so). We look at her dodgy doxastic history with indictments of her near and dear, and as a result attribute to her love or a desire for peace of mind. Of course, the latter does not outright qualify as an emotion. But it is motivated by love, or at least some care, for her family (otherwise we could not explain why her family's transgressions should disturb her peace of mind).

Theory, then, emerges as sufficient for emotion attributions both on the basis of behaviour and of knowledge of the other's mental states. In Part I, I promised to show Theory sufficient for *predictions* only once I had shown it sufficient for attributions. The reason was that at that stage I would have had to appeal to a phenomenological difference between predictions of unexperienced and of experienced emotions. It is time to fulfil my promise.

Section 3: Theory's sufficiency for emotion-predictions

Section 2 of this part suggested that a person can have the concept of an emotion without having experienced the emotion. It also showed that knowing the RS* of a particular emotion is sufficient for understanding what is involved in attributions. I then argued that such understanding, together with the appropriate generalisations, is sufficient for the ability to attribute the emotion in question. In this section I argue that the conclusion applies to predictions.

1. Predictions on the basis of knowledge of the other's mental states

The argument in section 2 focused mainly on attributions on the basis of behaviour. I did mention that attributions on the basis of other mental state is unproblematic because the cause of an emotion is partially constitutive of the RS*, and so knowing the subject's other mental states amounts to knowing the states antecedent to the emotion. Some general knowledge of paradigmatic narratives will be sufficient for picking out which of those mental states are causally relevant for the attributed emotion. But if we really know the subject's other mental states, or at least the relevant ones, then this point automatically applies to predictions. Certain mental states are more likely than others to bring about a particular emotion.

2. Predictions on the basis of behaviour and circumstances

This point can be extended to predictions on the basis of behaviour. Since knowledge of behaviour and circumstances is sufficient for attributions (section 2, sub-section 1.), having such knowledge gives access to the other's mental states. But once we have such knowledge (from the last argument), we are fully equipped for the requisite prediction.

3. Objections

It may be urged that I have just resurrected a serious problem I imputed to Heal. One of my arguments against Simulation's pretensions to heuristic parsimony was that Simulation assumed that two agents (the simulator and the simulated) would arrive at the same solution to a problem (Chapter 1, Part II, section 2). So, the objection would go, when I claim that knowing the mental states antecedent to an emotion is sufficient for a prediction of the emotion, I am claiming in effect that certain mental states would always give rise to the same emotion. This complaint neglects my argument from narrative in section 2 above. I argued there that the theorist, by contrast with the simulator, takes into account the person's character and knowledge of what he has previously done in similar situations and with similar ensuing behaviour. Simulation could not exploit such ethological knowledge because it was relying on a characterless and universal mechanism. By contrast, I have been urging, throughout my criticism of Simulation and my account of Theory, the necessity of ethological considerations. Moreover, in my analysis of emotions, I showed, by appeal to the constitutive role of narrative, that such considerations are internal to the concept of emotion, and so internal to an account of understanding what is involved in judgements about others' emotions. There is no obstacle, then, to acknowledging Theory's sufficiency for emotion-predictions.

Conclusion

Theory, then emerges as necessary and sufficient for emotion-predictions and attributions. The conclusion is rather embarrassingly obvious - after all, having the concept of any mental state entails that we know how to pick out its extension. But this is not the whole story. I said that knowing the revised Relational Schema, together with some generalisations about the standard links amongst its elements is necessary. These links were provided by talk of paradigmatic narrative structures. Let me sharpen the contours of this proposal. I adopted, in Chapter 1, Botterill's metaphor of a Lakatosian core, together with the principles he built into it. These were:

Action Principle: An agent will act in such a way as to satisfy, or at least increase the likelihood of satisfaction of, his/her current strongest desire in the light of his/ her beliefs.

Perception Principle: When an agent A attends to a situation S in a given way, and p is a fact about S perceptually salient in that way, then A acquires the belief that p.

Inference Principle: When an agent A acquires the belief that p and a rational thinker ought to infer q from the conjunction of p with other beliefs that A has, A comes to believe that q.

Considering our frequent appeal to emotions in explanations, it is clear that we need to supplement the core with the insights developed in this chapter. This, then, is the Theory which guides predictions and attributions of emotions:

(i) Knowing the RS* of a particular emotion gives us understanding of what is involved in judgements about others' emotions in general.

(ii) Knowledge of the paradigmatic narrative structures of (a) particular emotions and (b) character types, is sufficient for equipping us with paradigmatic relations between

(a) emotions and their causal as well as justificatory antecedents;

(b) character types and (a);

Knowledge of (i) and (ii), together with some knowledge of the character of the person predicted and his circumstances, behaviour and/ or the relevant mental states is sufficient for both predictions and attributions. This knowledge can be summed up as

Emotion Principle: When an agent A, who is a character type C and has narrative N, seriously attends to a situation S, A will typically experience emotion E.

The Theory which is necessary and sufficient for emotion-predictions and -attributions, then, is modeled on Botterill's analogy with a Lakatosian core, where:

(a) the central tenets are the Action Principle, the Perception Principle, the Inference Principle, and the Emotion Principle, which capture the conceptual links amongst mental states on the one hand, and amongst mental states, situational input, and action on the other;

(b) the periphery consists of general rules of thumb based on empirical knowledge about how things generally tend to work, which is partly based on how things generally tend to work *with me*;

(c) and the relevance of a particular set of precepts from this database is determined by personal salience, as well as a representation of how the ascriber compares with others.

CODA: Under the carpet

This dissertation has shown that Theory is necessary and sufficient for emotion-predictions and -attributions which cherish the 'from the inside idea'. I attempt here to bring to light some of the issues I have swept under the carpet, without making too much of a mess. I proceed as follows. First, I consider some of the ironic aspects of my conclusion. Second, I explore a possible dilemma which my conclusion can conjure. Third, I attempt, and fail, to apply the emotion-based arguments deployed in Chapter 3 to Theory's sufficiency for predictions and attributions of mental states *in general*. Finally, I consider emotions' role in our ethical lives.

1. There is a three-fold irony about my claim that Theory offers a more 'personal' way of attributing emotions.
 - a. First, Simulation was historically introduced in order to rectify Theory's functionalist, and so *impersonal*, approach to explaining others. I argued, however, that if Simulation postulates a mechanism, then it gets into worse trouble than my pet version of Theory. *Mechanisms* cannot accommodate reasons without prior understanding of the subject, because mechanisms are causal entities (Chapter 3, Part I, section 1). Moreover, even if they could accommodate reasons, we need to feed them with the relevantly similar, and so *full input*. This means that understanding of the full causal history of the emotion is necessary prior to the simulation procedure. Understanding of both the subject and the emotion's history brings the 'from the inside idea' into the picture, but neither kind of understanding can be accommodated by Simulation.
 - b. The second reason my conclusion is ironic is that Simulationists often rely on the central role that *imagination* and *empathy* allegedly play in predictions and attributions. And, of course, we all agree that these two capacities are the very emblem of the 'from the inside idea', so it is ironic that Simulation turns out to be inhospitable to this idea. The irony, I submit, is only apparent, because it stems from the ambivalent use Simulation makes of the notions of empathy, imagination and the 'from the inside idea'. First, I showed in my discussion of *Othello* that the kind of imagination relevant for simulation is not the noble faculty praised by poets (and which Othello has), but *empathetic* imagination (Chapter 1, Part III, section 3). Second, empathy itself was cashed out by Simulationists in terms of the ability to feed the appropriate mechanisms with 'pretend' input (Chapter 1, Part I, section 1). This then is how Simulationists capitalise on the ambivalence of imagination and empathy: When they say 'empathy', we think of the morally-laden capacity, closely associated with sympathy, and so

think of a first-personal approach to others; when the Simulationist says 'imagination', we think of the faculty praised by poets and involving the genuine ability of transcending ourselves. But what Simulationists *in fact* mean by the two terms is the ability to feed our actual mechanisms for emotions and practical and theoretical reasoning with 'pretend' input. And it is not clear that the input is relevantly similar to the mental states of the person whom we are trying to explain (as I argued in Chapter 3, Part II, section 1). Finally, Simulationists capitalise on the ambivalence of the 'from the inside idea'. When Heal introduces it, she talks about the subject's 'point of view', and so we automatically think of a genuine understanding of another. Accordingly, when she mentions imagination and empathy, we think of the nobler meanings of the terms. But, as should by now be obvious, all that the 'from the inside idea' means is the requirement that we allow room for reasons. Neither imagination nor empathy (in the noble or ignoble senses) is necessary for understanding that there are reasons for someone's being in a particular state. On the other hand, I suggested in my discussion of *Othello*, that the audience does accomplish a feat of imagination (in the noble sense), but it is in order to feel *sympathy*. Moreover, the relevant imaginative capacity is essentially Theory-driven.

c. This last point suggests a third way in which my conclusion is ironic, since 'Theory' has rather unimaginative and dry connotations. I hope to have diffused these overtones, by developing an account of Theory-driven judgements about others' emotions. Such judgements, I argued, involve *understanding* of the character of the person whom we are explaining, and so do not amount to some dry theoretical approach to others, devoid of human engagement in a world of other humans. This, then, is the sense in which Theory-driven imagination (in the nobler sense) is necessary for judgements of others' mental states. I never argued that imagination is exhausted by Theory (though the present reader's may well have been). Imagination still remains the faculty by which we transcend our own situation and character, and learn what is involved in the *experience* of unfamiliar emotions (though this is not necessarily how we learn the *concepts* of such emotions). The reason we need imagination to conceive of the *experience* of an unexperienced emotion is that feeling towards the emotion's object (which includes seriousness towards the construal and phenomenology) is partly *constitutive* of emotions. But it is clear that Theory will not help directly with phenomenology and seriousness. We need to *imagine* what it must be like to be in the grip of the unfamiliar emotion. To concede all this, however, is not to deny that Theory plays a fundamental role in imagination.

2. The second issue swept under the carpet is that my account may give rise to a dilemma⁹⁷: *either* (a) I have set up Simulation as a straw-man in order to show Theory necessary for predictions in general, *or* (b) I have captured Simulation faithfully but have smuggled it in my account of Theory in order for Theory's sufficiency to be plausible.

a. The first horn arises from noticing the obviousness of the claim that Theory is necessary for predictions. Surely, the objection would go, Simulation cannot really envisage predictions without any standards for relevant similarity between the ascriber and the subject? Two considerations should dispel such doubts. First, I have grappled in detail with a particular author's position, Heal's, and have quoted her extensively throughout Chapter 1. It is obvious from her position, as well as from what I have cited of Gordon's and Goldman's, that none of them thinks anything like Theory necessary for predictions. As seen in my account of Sunflower Simulation in Chapter 3, Goldman ignores such issues altogether; while Gordon thinks that they can be resolved by further Simulation (cf. Part II, section 1). Heal, as already noted (Chapter 1, Part II, section 3), does concede that Simulation "must allow somewhere for the idea of different personalities", but denies that such insight would derive from Theory. As also seen, she does not offer any suggestions as to where else such insights might be gleaned.

Second, the existence of so many Simulation-Theory hybrid positions attests to the faithfulness of my account of Simulation. What these eclectics add to Simulation *simpliciter* is precisely some Theory-laden psychological information (cf. Chapter 1, Part I, section 2). Of course, this is not the position I have developed here. The point rather is that referring to positions, which import Theory into Simulation, as *hybrid* positions, shows that Simulation does not cater for relevant similarity. This insight, however, gives rise to the second horn of the dilemma: have I not, despite my avowed intentions, smuggled Simulation in my account of Theory?

b. The second horn would be best understood if my argument for Theory's necessity for emotion-attributions is recalled. I urged there (Chapter 3, Part I, section 2) that a female viewer of *King Lear* must use Theory in order to ascribe paternal love to Lear and so predict that he would be devastated by Gonerill's and Regan's treatment of him, and grateful for Cordelia's love. The second horn arises as follows. First, we note that I have neglected an essential element of the prediction - the viewer's phenomenological experience during the prediction. We all have the intuition that the viewer is not merely wielding some lifeless theoretical tools, but is empathetically engaged in Lear's story. For my account to capture this phenomenological

⁹⁷ I am deeply indebted to Eusebius McKaiser for pointing out that my thesis is vulnerable to this line of attack, as well as for articulating the most pernicious version of this attack (with which I deal here).

aspect, the envisaged objection goes, I must either deny that we engage in such empathy, or acknowledge the empathy, and so smuggle Simulation's tools in the envisaged Theory-prediction. Obviously, I do not want to deny the phenomenological intuition, and so must concede that I have smuggled in Simulation, and so concede that my position is a (closet) hybrid one.

I have already mentioned that this sense of empathy is not what Simulation means by the notion. But suppose it was. Two considerations will avert the danger of the second horn. First, the argument about the female viewer was aimed at establishing Theory's *necessity* for emotion-attributions and not its sufficiency. In light of this, the claim of Theory's necessity is perfectly compatible with the insight that the female viewer may have been engaging in empathy (with all its relevant phenomenology). My claim that Theory is sufficient for attributions and predictions, however, is clearly incompatible with the phenomenological insight.

This is the second consideration, then: Recall that one of my arguments for Theory's sufficiency for emotions' attributions turned on Regan's, Gonerill's and Edmond's *lacking* the mechanisms (capacity for love) envisaged by Simulation and yet being *capable* of ascribing and predicting paternal love. The present point is that since they are incapable of love, they are incapable of the (love) phenomenology relevant for empathising with someone who loves another. This provides a clear example of someone who predicts and ascribes emotions without the phenomenology relevant for empathy (in the sense envisaged by the objection). It follows that although we may, and indeed do, engage in empathy, empathy, and so Simulation, is not necessary for emotion-predictions and -attributions. Perhaps Simulation is necessary for *empathy*, but this concession certainly does not commit me to a hybrid position in the context of predictions and attributions, since I have argued that empathy is not necessary for attributions and predictions. It further follows that I have smuggled Simulation (in the guise of this nobler sense of empathy) neither in my account of Theory, nor in my arguments for Theory's sufficiency for emotion-attributions and predictions. The dilemma, then, was spurious.

3. The third major issue swept under the carpet is this: considering that my arguments for sufficiency turned on the *idiosyncratic* nature of the objects of *emotions*, how can this account be extended to Theory's sufficiency for predictions and attributions of mental states *in general*? This is a serious problem for my account, and one which I am uncertain how to handle. I need to attempt it, though, because an account which is so narrow and not amenable to generalisation, cannot constitute a significant contribution to the Theory-Simulation debate.

I have gone some way towards handling the problem: I have shown Theory *necessary* for mental-state predictions on the basis of knowledge of the other's mental states. If we can show that such knowledge is necessary for predictions, then we can conclude that Theory is necessary for all mental-state predictions.

But we still need to show Theory *sufficient* for such predictions for symmetry's sake, and so that we can claim Theory's sufficiency for *attributions*. My intuition is that this would be impossible without a sophisticated ontology for each kind of mental state. Such ontologies are beyond the scope of this dissertation. But perhaps I have already indicated how such an ontology can be used to show Theory's sufficiency. Since I agreed with Roberts that all our thoughts consist of construals, we can use a similar argument to the one offered in Chapter 3 (Part II): construals are irreducibly personal; if they are the appropriate object of belief and desire (this is where it will get messy, I suspect), then everything said about emotions' objects would apply to them. This line of thought, unfortunately would merely confirm Theory's *necessity* for attributions and predictions. I take it that the sufficiency argument will have to consider someone who is incapable of certain thoughts, and yet can attribute/ predict them competently.

But this is where the argument hits a stumbling block and another issue swept under the carpet. There is a crucial asymmetry between emotions and beliefs⁹⁸. Beliefs' propositional content can be *doxastically* entertained without compelling the agent to believe the content or to act upon the belief, while emotions' objects cannot be *emotionally* entertained (though they can be doxastically entertained). While I can entertain a certain construal of a situation, the concerns which partly constitute the emotion's object will be more problematic. Of course, I can entertain occurrent desires and aversions, but if the concerns involve dispositions, or worse, some emotions, what does it mean for me to *pretend* that I have such emotions? This line of thought can be used in favour of Theory when emotions are at stake. It will, however, pose a great problem for showing Theory's sufficiency for attributions of other mental states, because we need a symmetry between emotions and other mental states if we are to extend the arguments for emotions: the negative arguments relied on a person's *incapacity* for certain emotions; but such incapacity is inconceivable in the case of beliefs whose propositional object can be entertained at will.

4. The final, and rather substantial, issue that I have swept under the carpet is ethics. Most studies of emotions are motivated by the crucial role emotions play in our moral and, more broadly, ethical lives. Some of the fundamental questions are: To what extent do emotions

⁹⁸ I am indebted to Marius Vermaak for this insight.

motivate action? How responsible are we for our emotions (considering their legendary status as passions)? A related issue is the role of emotions in our doxastic lives. Emotions have an egregious history for vitiating our 'judgement': they are traditionally construed as the chief culprits in *motivated believing* and *akrasia*. My circumvention of these issues is only partly due to cowardice. The main and quite deliberate reason for ignoring them is that too many ethics-based accounts derive the nature of emotions from their preferred conception of ethics. As a result, they either fail to capture faithfully emotions' central place in our lives and/ or reduce emotions to meaningless theoretical constructs. The unfaithfulness or reduction certainly make accounts of practical reasoning very comely, but the ensuing constructs are no longer recognisable as the palpable, complex phenomena which govern our lives in such important ways. One (the old) way of going about this disfigurement of emotions is by reducing them to passions which overcome us. The other, equally iniquitous way, is by reducing them to judgements. In light of these considerations, it is crucial that we have a complete picture of what sorts of mental states emotions are and how we handle them in explaining others' behaviour *prior to*, and as the *foundation* for, ethical commitments. Such a picture I claim to have offered.

My account has several implications for ethics, of which I offer but a sketch here (anything more will require a dissertation on its own). The gist of the implications is that my model renders certain types of excuses untenable.

a. First, if Simulation were right, we would expect that misattributing an emotion could be blamed on the (mis)attributor's having a limited capacity for certain emotions and so for their attributions. This would make for a great excuse⁹⁹ in cases where the misattribution leads to tragic consequences ('I just *couldn't* imagine that she would be so upset as to commit suicide because of what I told her'). The fact that we attribute and predict others' emotions by using Theory and knowledge of character, leaves no room for excusing emotional misunderstanding on the basis of faulty imagination, or incapacity for certain emotions. Explaining others' behaviour is not a matter of possessing the capacity for the relevant emotions, but of having an educated repertoire of emotion-concepts and their interrelations with other mental states and with behaviour. Concerning imagination, as already argued, it is itself grounded in Theory. The envisaged excuse would not work because the person should have developed their emotion-concept repertoire. The injunction for an educated repertoire could obviously not work for Simulation: compare 'You should have *known* better than to attribute indifference to Lear' with

⁹⁹ Thinking about how he used to ignore Lolita's need for a family, Humbert Humbert alludes precisely to this sort of excuse when he says: "I admit that a man of my power of *imagination* cannot *plead* personal ignorance of universal emotions" (2.32.287, my italics).

‘You should have developed your *capacity* for paternal love’. Moreover, as noted in my discussion of Tom Pinch, someone’s interpretation of another, his choice of a particular narrative palimpsest, can tell us much about the moral qualities of the interpreter (Chapter 2, Part III, section 3).

b. Second, the account disallows excusing emotions on the grounds either of non-rational influences (‘Oh, you must excuse my breaking your teeth in anger yesterday. I was drunk’), or of the emoter’s having been overcome by emotion (‘Oh, you must excuse his lashing out at you, he hasn’t had a grip on himself since his wife died’). Non-rational influences are at the heart of our embodiment, and as suggested in the case of Leontes, we must call his state an emotion despite the possibility that it may have been non-rationally induced, and still censure him for its subsequent unresponsiveness to reasons (that is for subsequent *irrationality*). As to being overcome by an emotion, I argued in Chapter 2 (Part I, section 1) that although one can be momentarily overcome, one can certainly cultivate long-term attitudes which would prevent one from being regularly overcome by certain emotions considered negative. (My account is, and will remain, silent about what may count as negative emotions.) This point follows from the role construals and concerns play in our emotional lives. An emotion is not just an irruptive state, visited on us by the gods; it has a history of concerns and construals. We can cultivate virtuous concerns (again my account offers no guidance as to what this might mean) and the tendency for accurate construals.

c. Third, the role emotions play in our doxastic lives is incipient in the last point. Emotions are traditionally construed as disruptive nuisances on the pristine horizon of our doxastic beatitude. But this is misguided. Recall that both thinking and emotions involve construals. The difference is that emotions are directed at a *concern*-imbued construal. This suggests that emotions are epistemically continuous with (though far more complex than) beliefs. The aim of the virtuous doxastic life, then, is not to eliminate emotions, but to harmonize the construals involved in emotions with those involved in beliefs.

The *claim* that my pet Theory is necessary and sufficient for emotion-attributions and -predictions, then, offers a solid foundation for moral and epistemic explorations of emotions. On the other hand, my *method* of arriving at the claim that Theory is necessary and sufficient, seems to debar us from similar claims about Theory’s role in predictions and attributions of *other* mental states. This is ironic in the sense that one would have expected Theory to be sufficient for the *less* complex mental states, since it deals in generalisations. The irony,

however, offers its own consolation: if Theory is sufficient for the *most complex* mental states, and is necessary for *all* mental states, showing it (albeit not by the method *I* develop) sufficient for *all* mental state attributions and predictions cannot be far around the corner.

BIBLIOGRAPHY

On the Simulation-Theory debate

- Blackburn, S. (1995) 'Theory, Observation, and Drama', in Davies and Stone (1995a).
- Bolton, D. (1995) 'Self-knowledge, Error and Disorder', in Davies and Stone (1995b).
- Botterill, G. (1996) 'Folk Psychology and Theoretical Status', in Carruthers and Smith (1996).
- Burge, T. (1996) 'Our Entitlement to Self-Knowledge'. *Proceedings of the Aristotelian Society*, 96 (91-116).
- Carruthers, P. and Smith, P. K. (eds) (1996) *Theories of Theories of Mind* (Cambridge: Cambridge University Press).
- Carruthers, P. (1996) 'Simulation and Self-Knowledge: a Defence of Theory-Theory', in Carruthers and Smith (1996).
- Currie, G. 'Imagination and Simulation: Aesthetics meets Cognitive Science' in Davies and Stone (1995b).
- Davies, M. and Stone, T. (eds) (1995a) *Folk Psychology: The Theory of Mind Debate* (Oxford: Blackwell).
- Davies, M. and Stone, D. (eds) (1995b) *Mental Simulation: Evaluations and Applications* (Oxford: Blackwell).
- Davies, M. and Stone, T. (1998) 'Folk Psychology and Mental Simulation', in O'Hear (1998).
- Dretske, F. (1988) *Explaining Behaviour* (Cambridge, MA: MIT Press).
- Davies, M. and Humphreys, G. W. (eds) (1993) *Consciousness* (Oxford: Blackwell).
- Fodor, J. (1995) 'A Theory of the Child's Theory of Mind', in Davies and Stone (1995b).
- Fuller, G. (1995) 'Simulation and Psychological Concepts', in Davies and Stone (1995b).
- Goldie, P. (1999) 'How We Think of Others' Emotions'. *Mind & Language*, Vol. 14 No. 4 (December 1999: 394-423).
- Goldman, A. (1995a) 'Interpretation Psychologized', in Davies and Stone (1995a).
- Goldman, A. (1995b) 'In Defense of the Simulation Theory', in Davies and Stone (1995a).
- Goldman, A. (1995c) 'Empathy, Mind, and Morals' in Davies and Stone (1995b).
- Gordon, R. M. (1995a) 'Folk Psychology as Simulation', in Davies and Stone (1995a).
- Gordon, R. M. (1995b) 'The Simulation Theory: Objections and Misconceptions', in Davies and Stone (1995a).
- Gordon, R. M. (1995c) 'Reply to Stich and Nichols', in Davies and Stone (1995a).
- Gordon, R. M. (1995d) 'Simulation Without Introspection or Inference from Me to You', in Davies and Stone (1995b).
- Gordon, R. M. (1996) 'Radical Simulationism', in Carruthers and Smith (1996).

- Heal, J. (1995a) 'Replication and Functionalism', in Davies and Stone (1995a).
- Heal, J. (1995b) 'How to Think About Thinking', in Davies and Stone (1995b).
- Heal, J. (1996) 'Simulation, Theory and Content', in Carruthers and Smith (1996).
- Heal, J. (1998) 'Understanding Other Minds from the Inside', in O'Hear (1998).
- Nichols, S., Stich, S., Leslie, A., and Klein, D. (1996) 'Varieties of Off-line Simulation', in Carruthers and Smith (1996).
- O'Hear, A. (ed) (1998) *Current Issues in Philosophy of Mind* (Cambridge: Cambridge University Press).
- Perner, J. (1996) 'Simulation as Explication of Prediction-Implicit Knowledge about the Mind: Arguments for a Simulation-Theory Mix', in Carruthers and Smith (1996).
- Sterelny, K. (2001) *The Evolution of Agency and Other Essays* (Cambridge: Cambridge University Press).
- Stich, S. and Nichols, S. (1995a) 'Folk Psychology: Simulation or Tacit Theory?' in Davies and Stone (1995a).
- Stich, S. and Nichols, S. (1995b) 'Second Thoughts on Simulation', in Davies and Stone (1995b).
- Stone, T. and Davies, M. (1996) 'The Mental Simulation Debate: a Progress Report', in Carruthers and Smith (1996).
- Stoneham, T. (2000) 'Self-Knowledge', *The Handbook of Epistemology*, (eds) Niiniluoto, I., Sintonen, M. and Wolenski, J. (Kluwer Academic Publishers).
- Wellman, H. M. (1990) *The Child's Theory of Mind* (Cambridge, MA: MIT Press).

On Emotions and Beliefs

- Bennett, J. (1990) 'Why is Belief Involuntary?' *Analysis* (March 1990: 87-107).
- Calhoun, C. and Solomon, R. (eds) (1984) *What is an Emotion: Classic Readings in Philosophical Psychology* (New York: Oxford University Press).
- Currie, G. (1998) 'Narrative'. *Routledge Encyclopedia of Philosophy*, vol. 6 (London: Routledge).
- de Sousa, R. (1980a) 'The Rationality of Emotions', in Rorty (1980).
- de Sousa, R. (1980b) 'Self-Deceptive Emotions', in Rorty (1980).
- de Sousa, R. (1987) *The Rationality of Emotion* (Cambridge, MA: MIT Press).
- Eilan, N. (1998) 'Perceptual Consciousness, Attention and Consciousness', in O'Hear (1998).
- Ekman, P. and Davidson, R. J. (eds) (1994) *The Nature of Emotion: Fundamental Questions* (New York, Oxford: Oxford University Press).

- Ekman, P. and Friesen, W. V. (1975) *Unmasking the Face* (Englewood Cliffs, NJ: Prentice-Hall).
- Elster, J. (1989) *Nuts and Bolts for the Social Sciences* (Cambridge: Cambridge University Press).
- Elster, J. (1999) *Alchemies of the Mind: Rationality and the Emotions* (Oxford: Oxford University Press).
- Gettier, E. L. (1962) 'Is Justified True Belief Knowledge?' *Analysis*, 23 (1962-1963: 121-3).
- Goldie, P. (2000) *The Emotions: A Philosophical Exploration* (Oxford: Oxford University Press).
- Goldie, P. (ed) (2002) *Understanding Emotions: Mind and Morals* (Aldershot: Ashgate).
- Goldie, P. (2002) 'Emotion, Personality and Simulation', in Goldie (2002).
- Greenspan, P. (1988) *Emotions and Reasons* (London: Routledge).
- Griffiths, P. E. (1997) *What Emotions Really Are* (Chicago: University of Chicago Press).
- Harre, R. and Parrott, W. G. (eds) (1996) *The Emotions: Social, Cultural and Biological Dimensions* (London: SAGE Publications).
- James, W. (1884) 'What is an Emotion?', *Mind* 9, (108-205).
- Kenny, A. (1963) *Action, Emotion and Will* (London: Routledge).
- MacLean, P. D. (1980) 'Sensory and Perceptive Factors in Emotional Functions of the Triune Brain', in Rorty (1980).
- Marshall, G. (1980) 'Overdetermination and the Emotions', in Rorty (1980).
- Montmarquet, J. (1993) *Epistemic Virtue and Doxastic Responsibility* (Lanham: Rowman & Littlefield Publishers).
- Morton, A. (1980) 'Character and the Emotions', in Rorty (1980).
- Rey, G. (1980) 'Functionalism and the Emotions' in Rorty (1980).
- Roberts, R. C. (1988) 'What an Emotion Is: a Sketch'. *Philosophical Studies* 77 (1988: 319-338).
- Roberts, R. C. (1995) 'Feeling One's Emotions and Knowing Oneself'. *Philosophical Review* 97 (2) (1995:183-209).
- Roberts, R. C. (1999) 'Emotions as Judgments'. *Philosophy and Phenomenological Research*, vol. LIX, No. 3 (September 1999).
- Rorty, A. O. (ed) (1980) *Explaining Emotions* (Berkeley: University of California Press).
- Rorty, A. O. (1980) 'Explaining Emotions', in Rorty (1980).
- Scruton, R. (1980) 'Emotion, Practical Knowledge and Common Culture', in Rorty (1980).
- Solomon, C. S. (1980) 'Emotions and Choice', in Rorty (1980).
- Stocker, M. (1980) 'Intellectual Desire, Emotion, and Action', in Rorty (1980).

- Stocker, M. (1983) 'Psychic Feelings: Their Importance and Irreducibility', *Australian Journal of Philosophy*, vol. 61 (5 - 26).
- Stocker, M. with Hegeman, E. (1996) *Valuing Emotions* (Cambridge: Cambridge University Press).
- Velleman, D. J. (2000) *The Possibility of Practical Reason* (Oxford: Oxford University Press).
- Williams, B. (1973) 'Deciding to Believe', in Williams, B. *Problems of the Self* (Cambridge: Cambridge University Press).

Literature

- Dickens, C. (1854) *Martin Chuzzlewit* (London: Penguin, 1994).
- Eliot, G. (1859) *Adam Bede* (London: Penguin, 1994).
- Nabokov, V. (1959) *Lolita* (London: Penguin Books, 1995).
- Shakespeare, W. *Othello*, in *William Shakespeare: The Complete Works*, edited by Alexander, P. (London and Glasgow: Collins, 1951).
- Shakespeare, W. *The Winter's Tale*. The Arden Shakespeare, edited by Pafford, J. H. P. (London: Methuen, 1966).
- Shakespeare, W. *King Lear*, edited by Halio, J. L. (Cambridge: Cambridge University Press, 1997).

