

Emotion and Moral Judgment; Moral Unanimity and Diversity

Mark MH Tan

A thesis submitted in fulfilment of the requirements for the degree of Doctor of
Philosophy in Philosophy, the University of Auckland, 2021.

Abstract

There are instances of moral unanimity. There are also instances of moral diversity. I argue that emotion-based accounts of moral judgment can plausibly explain such instances (and this counts as a reason in favour of such accounts). I also argue that such accounts can best a rival conventionalist account in at least one respect in explaining (instances of) moral unanimity and diversity. I make these arguments with the aim of prompting 'fence-sitters' to side with such emotion-based accounts of moral judgments.

Acknowledgements

To my wife, Wen Chin, who gave her all to keep my life and our little family together when I went about completing this thesis. Without her unfailing love, patience and sacrifice, this thesis would have been nothing but a pipe dream. To my mother (Madeline), brother (Evan), and uncle (David), their counsel and encouragement through the years have kept me going. I would not have made it this far without their constant support. To my supervisors, Glen Pettigrove and Tim Mulgan, who endured much “jabberwocky” when reading earlier drafts of my thesis. If there is a semblance of coherence in my thesis, I attribute it to their guidance. To them all, I am forever grateful.

Table of Contents

CHAPTER ONE	1
WHIRLWIND OF EMOTIONS	1
FEELING-BASED APPROACHES	6
THOUGHT-BASED APPROACHES	16
HYBRID APPROACHES	34
PRESUPPOSITIONS	41
CHAPTER TWO	44
EAMs	44
SOME EMPIRICAL SUPPORT	46
EMOTIVISM	52
EXPRESSIVISM	54
EMPATHY-BASED SENTIMENTALISM	61
DISPOSITIONALISM	64
PLURALISM	68
ADDRESSING SOME OTHER CONCERNS	74
CHAPTER THREE	82
MENCIUS, BUT WHY?	82
MAPPING ON TO WESTERN EMOTION THEORIES	83
FROM VIRTUES TO JUDGMENTS	88
MENCIAN PLURALISM	89
EMOTIVE SOURCE OF MORAL JUDGMENTS	92
THE NEXUS BETWEEN EMOTION AND BASELINE JUDGMENT	95
AN ALTERNATIVE SOURCE FOR BASELINE JUDGMENTS?	100
THE NEXUS BETWEEN EMOTION AND NON-BASELINE JUDGMENT	107
MENCIUS' COMMITMENT TO NON-BASELINE JUDGMENTS	109
BROADENING THE EPISTEMIC BASE	113
THE WATER AND SPROUT METAPHORS	119
INTERLUDE	126
CHAPTER FOUR	127
EXPLAINING MORAL UNANIMITY AND DIVERSITY	127
MORAL UNANIMITY	127

EXPLAINING MORAL UNANIMITY	130
MORAL DIVERSITY	140
EXPLAINING MORAL DIVERSITY.....	142
TUNING THE EMOTIONS.....	149
TUNING MORAL JUDGMENTS	163
TUNING THE EMOTIONS AND MORAL DIVERSITY	166
CONCLUDING REMARKS	171
CHAPTER FIVE.....	172
THE EXIT PLAN.....	172
ADJUDICATION PROBLEM.....	173
REPLY TO THE ADJUDICATION PROBLEM.....	180
SUMMARY	196
POSSIBLE ANSWERS THESIS.....	197
CONCLUDING REMARKS	201
CHAPTER SIX	203
A COMPARATIVE EXERCISE	203
(WHY) CONVENTIONALISM.....	204
CONVENTIONALISM’S EXPLANATION OF MORAL UNANIMITY	211
CONVENTIONALISM’S EXPLANATION OF MORAL DIVERSITY	217
LOSING OUT IN EXPLAINING MORAL UNANIMITY	222
FINAL THOUGHTS	236
BIBLIOGRAPHY	238

Chapter One

Whirlwind of Emotions

“[L]et us do away with the discrediting of the affective sphere and the heart. Let us expose the equivocation of the term “feeling” and clarify the different levels in this sphere” (Von Hildebrand 1965: 2007, 19).

“[W]e see nothing, just or unjust, which does not change in quality with a change in climate. Three degrees of latitude overthrow jurisprudence. A meridian determines that truth... It is an odd kind of justice to have a river for its boundary. Truth lies on this side of the Pyrenees, error on the other” (Pascal 1670: 1995, 23).

I start with an observation. Emotions pervade our *everyday life*. We laugh at jokes. We cry at funerals. We get disgusted at pungent smells and faeces. We get angry at someone who makes a rude remark. We fear the dark. We get anxious when our work is not going too well. We get all warm and fuzzy when we watch touching movie scenes. Emotions also play an important part in our everyday life. Metaphorically, they put colour into our lives. Without them, our lives would be colourless. Often, we only realise the importance of emotions in the counterfactual of their absence. I am reminded of a novel written by Lois Lowry, *The Giver*, of a dystopian world perceived only in black and white because of the suppression and eradication of emotions. Everything in this world seemed so ‘cold’ and ‘heartless’.

I make another observation. Emotions pervade our *moral lives*. We react with disgust at incest or cannibalism. We get outraged when we hear of some horrendous crime on the news. We admire soldiers who lay down their lives for their countrymen. We feel pity for the poor. It is also obvious that emotions such as these play an important part in our moral lives. Again, let us consider a counterfactual scenario. Imagine a world where there is no correlation between emotion and

morality. Imagine emotional indifference in the face of what we now regard as horrendous crimes, heroism, suffering, etc. It would at the very least appear odd and unfamiliar to many. For some, the counterfactual may lead them to consider whether emotions are (in one way or another) necessary for morality. Emotion-based approaches to morality (EAMs) take this consideration seriously. They argue emotions are (in one way or another) essential to moral judgments.¹

Before I lay out my plan for this chapter, here is another observation. There are instances of *moral unanimity and diversity*. On one hand, there are swaths of the moral domain where there is consensus across cultures. There is unanimity on an abstract level, where all cultures regard it as morally obligatory to care for others; or where all cultures find it morally obligatory to give unto others what they deserve. There is also unanimity on a more concrete level, where all cultures (in one way or another) find it morally wrong to kill an innocent person gratuitously; or where all cultures find treacherous behaviour to be morally reprehensible. On the other hand, there are pockets (at times large ones) of moral differences among cultures. There may be moral differences on an abstract level, where cultures seemingly have conflicting values. There are also moral differences on a more concrete level. We can see one in the *Analects*: ‘The Duke of She said to Confucius, “Among my people there is one we call ‘Upright Gong.’ When his father stole a sheep, he reported him to the authorities.” Confucius replied, “Among my people, those who we consider ‘upright’ are different from this: fathers cover up for their sons, and sons cover up for their fathers. ‘Uprightness’ is to be found in this”’ (Confucius c 500 BC: 2003, 13.18). In this passage we see that there is a distinction between *your* people and *my* people, your culture and mine. And we see how (at least from the outset) both cultures affirm the moral requirement to be ‘upright’. But despite this common affirmation, they differ on what we ought to do in concrete circumstances. The Duke of She seems to think that an upright person ought to report his father to the authorities if he has stolen a sheep, but Confucius says otherwise. He asserts that an upright man ought to conceal the faults of his father. I should be clear here that I am not so much concerned about which conception of uprightness is correct or which values are the right ones (if any). I make the above observation to segue into the question *can EAMs explain moral unanimity and diversity*. The key aim of my thesis is to answer this question. I would argue that EAMs can explain moral unanimity and diversity (and this counts as a reason in support of EAMs). I would also argue that EAMs do so in a manner

¹ I will discuss this in detail in the next chapter.

that is better than one rival account in at least one respect.² Hopefully, this would prompt fence-sitters to side with EAMs.

I would start by laying out in this chapter emotion theories. This is an important first step as there are conflicting answers to the question *what is an emotion*. In this chapter, I aim to address these conflicting views. I would discuss only some prominent theories as the ultimate aim of my thesis is not to provide a fully comprehensive description of emotion theories but to provide sufficient information on emotion theories to show how EAMs can explain moral unanimity and diversity. Though I would highlight some objections against the emotion theories discussed and potential replies, I do not aim to settle any debates between emotion theories. Instead, the objections and replies discussed aim to draw out and make clear certain presuppositions that I would carry into subsequent chapters of my thesis. They are:

1. Emotions are more than just feelings.
2. Emotions are distinguishable from each other.
3. Emotions are meaningful.
4. Emotions include not only adult human emotions but infant and animal emotions.
5. Emotions are malleable by thoughts, practices, norms, etc.

I would discuss the above with more detail later in this chapter. At this point, I provide the reader with an example of why these presuppositions are important for my thesis. In terms of the first presupposition, it arises out of a common objection against emotion theorists who argue that emotions are essentially feelings, i.e. emotions are not just brute feelings. Though I would consider potential replies to this objection, the content of this objection is a presupposition that I would take into subsequent chapters. This presupposition is important as it serves to preclude potential EAM sceptics who may argue that brute feelings constitute moral judgments. I would discuss this and other presuppositions in detail later and it would suffice to note at this stage that they serve to make clear my ecumenical approach toward emotion theories and highlight attributes of emotions that enhance the plausibility of EAMs in explaining moral unanimity and diversity.

Let us now start with emotion theories. To briefly recap, the chief claim of EAMs is that emotions (in one way or another) are essential to moral judgments. The concern of EAMs lies primarily in

² More on this in the final chapter.

elucidating the nexus between emotions and moral judgments. But I would miss a step if I jump directly into examining this connection. It is important to first have some understanding of the emotions before examining their connection to moral judgments. Unfortunately, 'emotion' is an elusive concept. For starters, emotion terms can also be virtue/vice terms (Morton 2009). Compassion can be an emotion (we can react with compassion for the beggar down the street). Compassion can also be a virtue (Jesus, for example, is a man of compassion; he has a stable disposition to act with compassion). Anger can be an emotion (I can react with anger when accused of a crime I did not commit). It can also be a vice (I can be an angry man; I have a stable disposition to act with anger). In our quest to better understand emotions, it would do us good by distinguishing emotions from such character traits.

But even if we distinguish emotions from character traits, we can also see that emotions take up various forms. I already alluded to two forms of emotion. An emotion can be occurrent (I can react with fear at a snake that slithered past me). An emotion can also be dispositional (when someone says that she fears snakes, what she might mean is that she is disposed to react with fear if a snake appears). At the point when the snake slithers in front of me, I may experience fear but not be aware of it. It might all have happened too quickly and I might only have realised that I had a reaction of fear while I was running for my life. Thus, I can be (un)aware of my emotional experiences. This fear of snakes seems less cognitively loaded when compared to the fear of not completing my PhD. It seems that little or no thought is involved in my fear of snakes. But surely some thought is involved in my fear of not completing my PhD. Hence, emotions can be on a spectrum of cognition (with little cognitive content on one end and being cognitively loaded on the other). Notice however that I have only considered distinct forms of emotions and have not addressed what they actually are.

What then are emotions? This is a hard question to answer. To explain my point, let us take my fear of not completing my PhD as an example. First, this fear is something I *feel*. I feel sick (a physiological experience). I feel a vague sense of anxiety (a phenomenological experience). Perhaps we can say that feelings are constitutive of my fear. Second, this fear involves thought. I think about the 'dangerous' consequences of presenting a not-so-competent piece of work or failing my oral defence. I think about the 'dangers' of what others would think of me. Would

people brand me as a failure? Considering these thoughts, perhaps we can also say that thoughts are constitutive of my fear. Third, this fear prompts me to act. At times, I am tempted by fear to give up, to just go on with my life and forget academia altogether. Most of the time, however, fear prompts me to study harder. I probably would not be an Immanuel Kant or David Hume, but I sure can produce a competent piece of work if I continuously put my effort into it. Now with this perhaps we can say as well that my fear has a motivational component. Now why do I feel sick or anxious? Why am I motivated to study harder? Perhaps it is because I appraise the non-completion of my PhD as something dangerous. If it happens, it might derail my aspiration to be an academic. I cannot then live my dream. This leads me to the fourth component. Perhaps we can say that my fear is also constituted by an appraisal of some sort. The next question we can ask is are all or any of these constituent parts essential to an emotion. Christine Tappolet makes clear this question in the following way: “Standard episodes of emotions involve a number of distinct components, such as what is often called a ‘cognitive basis’ – you need to perceive the drop of the plan, or to see your friend, in order to experience any emotional reaction – as well as a kind of possibly unarticulated appraisal, physiological changes, facial expressions, characteristic feelings, cognitive and attentional processes, and finally, a motivational component. The question then, is which, if any, of these components are essential” (Tappolet 2018, 495). Emotion theories attempt to answer this question. And they can be broadly placed into three camps: (1) feeling-based approaches (FBAs), (2) thought-based approaches (TBAs), and hybrid approaches (HYAs).³

³ I stress that this is not the only way of carving up the territory. Andrea Scarantino and Ronald de Sousa, for example, splits the territory into four traditions (feeling, evaluative, hybrid, and motivational traditions) (Scarantino and de Sousa 2018). I did not carve out the territory in the manner similar to Scarantino and de Sousa as my goal of outlining emotion theories (as discussed above) is to summarize emotion theories that would facilitate my discussion on how EAMs can explain moral unanimity and diversity. I do so by elucidating the emotion theories in a manner that ‘fleshes out’ the presuppositions that I would carry into the subsequent chapters. I would not be discussing the motivational tradition not because it is unimportant but because we do not require the services of motivational theories (though I would refer to some of their key features in subsequent chapters) to outline how EAMs can explain moral unanimity and diversity. The focus of my thesis is not on the link between emotion and moral *action* (if this is the case, then the motivational tradition would be key), but between emotion and moral *judgments*. John Deigh also carves the territory up differently. He argues that “[t]wo major themes characterize the study of emotions in modern philosophy and psychology. One is the identification of emotions with feelings. The other is the treatment of emotions as intentional states of mind, that is, states of mind that are directed at or toward some object” (Deigh 2009, 17). Unlike Scarantino and de Sousa, Deigh splits the territory into two. As you would have noticed, I split the territory into three instead. I did so mainly to clarify that there are HYAs that somehow attempt to capture both the feelings and intentionality aspects in their accounts of emotion.

Feeling-based Approaches

For FBA theorists, they argue that emotions are necessarily constituted by feelings. Let us start with William James as he argues that “our feeling of [bodily] changes as they occur IS the emotion” (W. James 1884, 189-190). Emotions are *feelings* produced by bodily changes. Feeling, for James, is necessarily constitutive of an emotion. As Jesse Prinz puts it for James, “[r]age is a perception that includes muscles tensing up and blood running to the extremities; fear is a perception that includes muscles freezing, respiration becoming constricted, and hairs standing on end” (Prinz 2008, 708). To support his FBA, James argues that: “[i]f we fancy some strong emotion, and then try to abstract from our consciousness of it all the feelings of its characteristic bodily symptoms, we find we have nothing left behind, no ‘mind-stuff’ out of which the emotion can be constituted, and that a cold and neutral state of intellectual perception is all that remains’ (W. James 1884, 193). If we remove the phenomenology of feelings, it seems very counterintuitive to say that one is still having an emotion. It seems oxymoronic to say (for example) that I have compassion for the suffering of that homeless person down the street but yet feel nothing. For James, bodily feelings are necessary for any emotional experience. This point largely extends to other FBAs. Robert Solomon’s nutshell statement of FBAs is aptly useful here: “An emotion might be ‘triggered’ by a disturbing perception, but its essence was physiology plus feeling” (Solomon 2004, 134). At this point, I note that any such theory that prioritizes bodily feelings seems to have an advantage. FBAs advocate for a common-sense approach. Emotions so often (if not always) correlate with feelings. Thus, it appears to make a lot of sense to say that emotions are constituted by such feelings. As Solomon argues, “... accounting for the bodily feelings (not just sensations) in emotion is not a secondary concern and not independent of appreciating the essential role of the body in emotional experience” (Solomon 2004, 149-150). And any theory aligned with common sense would, *ceteris paribus*, face fewer objections.

James’ FBA may be aligned with our intuition on the front of emotions being essentially feelings. But there seems to be an intuitive *misalignment* when we consider his understanding of feelings. The feelings involved here appear brute in the sense that they appear to be directed at *internal* physiological changes (as opposed to *external* events of the world). As John Dewey argues: “‘Common sense’ and psychological sense revolt at the supposed implication that the emotional ‘feel’ which constitutes so much of the meaning of our lives is a chance arrival, or a chance super-

imposition from certain organic changes which happen to be going on" (Dewey 1895, 16). One could respond to this counterintuitive objection by not taking it too seriously. One can reply that James knows acutely that his view is counterintuitive. He knows that "[o]ur natural way of thinking about... emotions is that the mental perception of some fact excites the mental affection called the emotion, and that this latter state of mind gives rise to the bodily expression" (W. James 1884, 189). He also goes on to argue that "we feel sorry because we cry, angry because we strike, afraid because we tremble, and [it is] not that we cry, strike, or tremble, because we are sorry, angry or fearful, as the case may be" (W. James 1884, 190). Hence, for James, the sequence of an emotional event is not:

Subject perceives a tiger → Subject experiences fear → Subject flees from the tiger

Rather, it is:

Subject perceives a tiger → Subject flees from the tiger → Subject experiences fear

This is even more counterintuitive, and he knows this. One response to Dewey is that: *yes, it is counterintuitive, but so what?* Another response to the counterintuitive objection is to disassociate what is counterintuitive from what is wrong. It seems rather counterintuitive (at least some time ago), for example, to think of earth as a sphere. But this view is not wrong. My point here is not that James' view is correct but that the objection that his view is counterintuitive carries negligible weight. This segues us into a more pressing objection against James' FBA. As Dewey argues, "[h]ope, fear, delight, sorrow, terror, love, are too important and too relevant in our lives to be in the main the 'feel' of bodily attitudes which have themselves no meaning" (Dewey 1894, 563). Dewey is arguing that emotions are so important in our lives, they cannot be just feelings about our physiological changes. Essentially, the objection is that there must be something more to emotions than feelings.

I would focus on two strands of thought regarding this 'there is more to emotions' issue. The first strand involves the meaningfulness problem. I will start with this. Andrea Scarantino and Ronald de Sousa explain Dewey's dissatisfaction this way: "Dewey was unhappy with the reversal of common sense entailed by the Jamesian idea that emotions are feelings that emerge in response to proprioceptions. If we truly were angry because we strike, Dewey countered, anger could not cause the striking, and this would deprive anger, as well as other emotions, of their explanatory

importance" (Scarantino and de Sousa 2018). Now perhaps we can concede that James' FBA cannot adequately respond to this meaningfulness problem, where emotions appear to be just brute feelings. However, other FBAs inspired by James' approach may be able to do so. For Prinz, emotions despite being constituted by feelings of physiological changes can still retain their meaningfulness. Emotions can still explain why we do what we do. Prinz argues that emotions are embodied appraisals in that they are "states that appraise by registering bodily changes" (Prinz 2004, 78). Mental episodes like the perception of a snake or a judgment that I am in danger can cause physiological changes. Fear then is the perception of such bodily changes. However, for Prinz, this fear is not just that. It is also an appraisal as it represents "changes in organism-environment relations" (Prinz 2004, 78). Prinz continues by arguing that "[l]oosely speaking, palpitations serve as evaluations... Feelings can obviate the need for cognition, because feelings carry information. The discrete motions of our bodies convey how we are faring in the world" (Prinz 2004, 78). In short, emotions still carry explanatory importance. They are in themselves meaningful. For some, this would be a satisfactory move as it addresses both how emotions can be essentially feelings but yet keep their meaningfulness. But others may think Prinz's response somewhat falls short as what he offers is only a third-person perspective of the meaningfulness of emotions. For dissenters, they can argue that the meaningfulness of emotions must be of the first-person, phenomenological sort. Emotions must not just be meaningful. They must also be meaningful *to me* when I experience them.

The second strand of thought that emerges from Dewey's 'there is more to emotion' claim involves the intentionality problem. For example, emotions appear to be intentional. They are directed at external objects. They are about something. My sadness is directed at the loss of my favourite Star Wars T-shirt. My fear is directed at a tiger that is growling right in front of me. My anger is directed at someone who punched me in the face. Franz Brentano offers an acute description of this intentionality aspect by arguing that "[e]very mental phenomenon includes something as object within itself, although they do not all do so in the same way. In presentation something is presented, in judgement something is affirmed or denied, in love loved, in hate hated, in desire desired and so on" (Brentano 1874: 1995, 68). Considering this, the FBA objector might argue that if emotions are essentially feelings about physiological changes, their intentional aspect is sidelined. Tiredness (e.g.) is a feeling about whatever physiological state that is associated with it. But it is not directed at anything. It has no intentionality. Dietrich Von Hildebrand uses the term 'mere

states' to explain such feelings: "[m]ere states exist in us as simple facts; their objective causal determination presupposes no knowledge of it; but even when knowledge is given, these states are not connected with their cause in a conscious meaningful way" (Von Hildebrand 1953, 173). In other words, James' claim that emotions are feelings of bodily changes seems to lack this important aspect of intentionality. What James would call emotions would be what Von Hildebrand would call mere states.

Now how do other FBAs address the intentionality problem? Antonio Damasio tries to do so while holding on to the claim that emotions are essentially feelings. He argues that "[i]f an emotion is a collection of corresponding bodily changes in body state connected to particular mental images that have activated a specific brain system, *the essence of feeling an emotion is the experience of such changes in juxtaposition to the mental images that initiated the cycle*" (Damasio 1995, 145). There are two perceptions involved here. First, the perception(s) of the event that triggered the emotion. Second, the perception of our internal bodily states. The former superimposed on the latter (Damasio 1995, 146) and what Damasio means by a feeling is the combination of both these perceptions. As we can see, the intentionality problem is then apparently solved by distinguishing between different types of feelings. The feelings of an emotion are similar to feelings of tiredness in that both are directed at one's bodily states. But the feelings involved in emotions are juxtaposed by the perception(s) of the event that triggered the entire process: "that experience of what your body is doing *while* thoughts about specific contents roll by, is the essence of what I call a feeling" (Damasio 1995, 145). This is where the feelings of an emotion are distinguished from feelings of (e.g.) tiredness. In terms that would address the intentionality problem, feelings of an emotion are directed at *internal* bodily states while thoughts of *external* events occur. It is in this sense that emotions are intentional.

However, some may argue that Damasio does not solve the intentionality problem. Instead, he evades it by redefining what intentionality means. John Deigh puts it this way by arguing that "... while [Damasio's] theory includes a thesis attributing intentionality to the feelings of emotion, the intentionality of these feelings, as Damasio understands it, does not correspond to the intentionality of the emotion that gives rise to it" (Deigh 2009, 34). When we say that emotions have intentional objects, we usually mean that they are directed at external objects/events

simpliciter. We do not mean that they are directed at internal bodily states plus thoughts of external objects. It seems awkward to say that my feelings of anger are directed at my rising heart rate, my blood boiling and external objects. But perhaps awkwardness would not sound the death knell of Damasio's FBA. We can apply here the same response to the counterintuitive objection. Just because it is awkward, it does not mean that it is wrong. Experiments on the movement of particles on a quantum level yield awkward results, but it does not mean that they are wrong. However, this awkwardness would be a problem if the view that internal bodily states plus thoughts of external objects are the intentional objects of emotions is incorrect. To some, such a view is not satisfactory as it is wrong. The phenomenology of the intentionality of emotion does not square with the view that emotions are directed at internal bodily states. As Deigh argues:

“[w]hen you fear an assailant who is threatening you with a knife, for instance, the object of your fear is the assailant. In experiencing such fear, your heart may race and your breathing may be labored. As a result, the experience may include your feeling your heart racing and your being short of breath. But these feelings, though part of your experience of fear, don't change what the object of your fear is. It is still the assailant, in this case, and not your racing heart or your shortness of breath” (Deigh 2009, 35).

Prinz's FBA offers an alternative solution to the intentionality problem. Like Damasio, Prinz argues that emotions are “states that register bodily changes” (Prinz 2004, 58). But unlike Damasio, Prinz rejects that “perceptions of bodily changes [are] *coupled* with evaluations” (Prinz 2004, 59). Instead, Prinz argues that perceptions of bodily changes *represent* “organism-environment relations with respect to wellbeing” (Prinz 2004, 52). Prinz acknowledges the flaw in James' FBA. He argues that “James was right that emotions are perceptions of bodily changes, but he failed to notice that such changes can represent exactly what judgments (cognitive appraisals) represent” (Prinz 2008, 209). Prinz then attempts to establish a clearer nexus between the internal state of a subject experiencing an emotion and her relationship with the outside world. My sadness is *directed* at the loss of my job in that it *represents* the loss involved. My anger is *directed* at my employer who made empty promises in that it *represents* the offence caused. However, one may still think this manoeuvre is not entirely satisfactory. Damasio redefined intentionality by changing what emotions are directed at. Prinz makes a somewhat similar move. But instead of changing what the emotion is directed at, he changes the meaning of 'directed at' itself. To say that sadness is *directed at* loss is to say that it *represents* loss. Directedness is *replaced* with representation.

This replacement evades the intentionality problem. It is easy to see that the claim that my anger is directed at the idiot who punched me should be distinguished from the claim that my anger represents the idiot's offence and how that offence relates to my experience of internal bodily states. To say that emotions are directed at objects is to offer a first-person or phenomenological account of the relationship between an emotion-eliciting event and the emotion itself. But to say that emotions represent organism-environment relations offers a third-person or impersonal account of that relationship. As Deigh argues in relation to the emotion of fear and neo-Jamesian accounts of emotion, "[n]othing in this relation of representation, however, implies that these feelings are directed at or toward any object in the world, real or imagined. Nothing in the relation implies that they orient you toward the threat you are facing" (Deigh 2009, 36).

FBA theorists who argue for the essentiality of noncognitive feelings in emotions also face another problem. If emotions are essentially feelings about physiological changes, it is not clear how such FBAs could distinguish between the varieties of emotion. The Schacter-Singer experiment can elucidate this distinction problem (Schacter and Singer 1962). In this experiment, subjects were injected with epinephrine and were asked about how they felt. The results revealed that their answers (either anger or euphoria) depended on the situations they were in. This suggests that anger and euphoria have the same profile of bodily changes. If the suggestion is correct, then FBAs may not be able to distinguish between anger and euphoria based on their profile of physiological changes. This experiment may seem to doom many FBAs. But upon closer scrutiny, it may fall short in falsifying the FBA claim that emotions can be distinguished based on their distinctive bodily profile.

Prinz argues that the Schacter-Singer experiment makes two assumptions (Prinz 2004, 70-72):

- A1: When subjects exhibit different emotional behaviour, it means that they are experiencing different emotional states.
- A2: Subjects are exhibiting different emotional behaviour despite having the same bodily states.

For Prinz, we can challenge both assumptions. For A1, subjects may exhibit signs of anger while feeling euphoric internally as a result of the epinephrine injection. For A2, the epinephrine injection may serve as a catalyst that induced the different emotional responses. Once the emotion occurs, the bodily changes that underlie it are not necessarily the same. Prinz continues to argue that even if the assumptions are correct, the objection still fails. Emotion 1 may be associated with a bodily profile of general arousal plus a distinctive set of physiological changes (P1). Emotion 2 with general arousal plus another distinctive set of physiological changes (P2). Generally, the brain state of Emotion 1 (B1) is triggered by general arousal plus P1 and Emotion 2 (B2) by general arousal plus P2. If this view is correct, B1 and B2 can occur without P1 and P2. This means that the general arousal induced by epinephrine can trigger B1 and B2 without P1 and P2. As such, the results of the Schacter-Singer experiment need not be interpreted as an affirmation of the claim that different emotions have the same bodily profile.

Damasio's argument is similar to that of Prinz. For the former, though emotion when experienced involves an interaction between the brain and body, there are neural devices that allow emotion to be experienced without this interaction. Emotion can occur in an 'as if' loop (Damasio 1995, 155). Brain states are caused by bodily states in general. But brain states once caused by such bodily states, through repeated associations of the external event and bodily states, no longer need to be triggered by actual bodily states. A semblance of feeling can be 'conjured' by the brain itself after perceiving an external event. The process of physiological changes is 'skipped'. In other words, bodily changes need not occur for one to have an emotional experience. As such, even though the subjects experience only general arousal due to the epinephrine injections, and there are no changes associated with the bodily profiles of anger and euphoria, it does not follow necessarily that anger and euphoria have the same bodily profile. They may just be experiencing the emotions of anger and euphoria in an 'as if' loop without the occurrence of the distinctive physiological features associated with anger and euphoria.

Now it looks like Prinz and Damasio have averted the Schacter-Singer problem. This being said, though the experiment does not successfully undermine FBAs, a sceptic may remain doubtful as it still seems far-fetched that each emotion has a distinctive bodily profile. After all, the *prima facie* list of emotions is long, and it is unlikely that every emotion on that list has a distinctive bodily

profile. Martha Nussbaum argues that emotions do not have distinctive bodily profiles and warns against any attempt to distinguish emotions based on their bodily profiles: “What we should not do is to associate a given emotion type with any one particular sort of feeling state” (Nussbaum 2016, 252). Compassion, for example, is what many would regard as an emotion. But it does not seem to be associated with a distinctive set of physiological changes. Love is another example. Many regard this as an emotion. And it can be connected to many distinctive physiological changes (a racing heart, flushed cheeks, sweaty palms) or to none at all.

FBA's may want to resolve this distinction problem by making a more modest claim. Instead of claiming that *all* emotions have distinctive bodily profiles, an FBA theorist may claim that only *basic* emotions have them. The plausibility of each emotion having a distinctive profile is higher if the list is shorter. Paul Ekman and Daniel Cordaro argue that there are seven basic emotions: anger, fear, surprise, sadness, disgust, contempt and happiness (Ekman and Cordaro 2011, 365). Though Ekman distinguishes these emotions mainly by their facial expressions, FBA's can use his list to identify distinctive bodily profiles for the basic emotions. This, however, seems to come with a steep theoretical cost. The implication here seems to be that non-basic emotions are not strictly emotions. This worry may be somewhat assuaged by what Ekman and Cordaro mean by a basic emotion. Ekman and Cordaro argue that each basic emotion is not an emotion per se but an *emotion family* (Ekman and Cordaro 2011, 364-365). For example, within the family of the basic emotion of anger, there is outrage, irritability, etc. Each emotion family has a theme that is phylogenetically developed (the product of evolution). It is this theme that unites them as a family. And what distinguishes the variants within an emotion family are their ontogenetic variances. As such, the proposal above is not for FBA's to focus on only seven emotions (to avert the distinction problem) but on seven emotion families.

Furthermore, FBA's need not rely on bodily profiles alone to distinguish between emotions. Perhaps Ekman and Cordaro's description of a basic emotion reveals another criterion that would allow FBA's to distinguish between emotions. Apart from arguing that basic emotions have a “distinctive physiology”, they also argue that such emotions have “distinctive universal signals” (Ekman and Cordaro 2011, 365). I singled out this feature because taking into account such ‘signals’ does not appear to be at odds with the key FBA claim that emotions are essentially feelings. Some

emotions (assuming that they are genuine) can be distinguished based on their facial expressions. Most of us can, for example, easily distinguish between anger and sadness. If so, this leaves FBAs open to claim that at least some emotions can be distinguished by their bodily profiles and their facial expressions. If we cannot distinguish anger from euphoria based on bodily profiles alone, we can probably distinguish them based on their facial expressions.

One problem however is that to claim that some emotions have 'distinctive universal signals' assumes that there is cross-cultural agreement in one's matching of facial expressions to such emotion terms (and this assumption is unwarranted). Jerome Neu challenges Ekman's assumption (Neu 2000, 20-22). Neu argues that the results of Ekman's experiments (showing how culturally isolated people can associate certain facial expressions with certain emotion terms in a manner similar to other cultures) do not necessarily support the claim that there is cross-cultural agreement on emotion terms and corresponding facial expressions. First, the Fore in New Guinea do not distinguish between surprise and fear expressions. Second, the subjects were not free to provide their own interpretation of the facial expressions. They were required to select from a list of emotions. It is possible that if given the opportunity to provide their own interpretation, they may associate the facial expressions with emotions not on the list. Third, the experiments were based on the presumption that we can recognize an emotion via its facial expression without its context. This presumption is not always correct. Facial expressions by themselves are often inadequate in the identification and discrimination of emotions. The context "as understood by the person having the emotion" helps (Neu 2000, 21). The facial expression usually associated with sadness (crying, for example) may be an expression of joy. The tears may be tears of joy.

Now even if FBAs adequately respond to Neu's objection, and they can show that there are distinctive universal signals for basic emotions, one may object that many other emotions lack such signals. Jealousy and love (for example) do not seem to have universal signals. As Charles Darwin aptly points out, "[p]ainters can hardly portray suspicion, jealousy, envy, etc., except by the aid of accessories which tell the tale" (Darwin 1872: 1965, 79). Relying on facial expressions and bodily profiles alone may only allow an FBA to distinguish between some emotions, but not all. One can still insist that large swaths of emotions are excluded by such a limited approach. It is not clear, for example, where *schadenfreude* fits in these emotion families. Nonetheless, starting with basic

emotions may allow FBAs to weave a plausible narrative of how emotions (despite their variety) can be distinguished from one another. Prinz argues that “[i]f we begin with a small stock of species-typical, basic emotions, we can expand our repertoire in two ways. First, we can combine these basic emotions together (thrills may arise when joy and fear blend), and second we can take a basic emotion or blend of emotions and ‘recalibrate’ it to respond to a new class of eliciting events... a small stock of emotions can be extended to represent new contents, and when that happens, new emotions are born” (Prinz 2008, 710). As we can see, for Prinz, the list of emotions is long. We start with the basic emotions, but they can be blended. A blend of emotion A and B can produce emotion C. A blend of emotion C with D might produce E. And so on. A, B, C, D and E can be recalibrated to have different eliciting conditions and produce other emotions. The point here is that Prinz’s FBA encompasses a wide range of what we would commonly regard as emotion. This range includes *schadenfreude*. For him, *schadenfreude* is “a feeling of joy that has been triggered by a cognition (or perception) of someone suffering” (Prinz 2004, 147). It is essentially calibrated joy. *Schadenfreude* is calibrated joy in the sense that it is triggered not by the fortunes of others (of which is paradigmatic of joy) but by the misfortune of others. This then opens the door for Prinz to argue that emotions are “individuated by perceived bodily patterns *and* by what they represent. The representational content depends on what a given perceived bodily state has the function of being reliably caused by” (Prinz 2008, 710). This would mean that even if joy and *schadenfreude* have the same bodily profiles, they can still be distinguished by what they represent. *Schadenfreude*’s representational content is distinct from joy’s in that the former specifically represents the suffering of others. Prinz’s response may also apply to basic emotions that appear to have indistinguishable bodily profiles. Walter Cannon argues (for example) that anger and fear involve visceral reactions that are indistinguishable from each other and hence the reactions cannot be used to distinguish between these emotions (Cannon 1929). Even if this is so, it is clear that anger and fear have different representational content and can as such be distinguished on this basis.

At this point, I remind the reader that the above objections serve primarily to draw out the presuppositions I would carry into the subsequent chapters (and I would make these presuppositions clear after considering some objections against TBAs). And the replies are not aimed at providing a complete defence of the objections to any emotion theory. Rather, they are aimed at (a) providing a response that would hopefully show the reader that a particular emotion

theory remains plausible and (b) serving as a segue into a discussion on other emotions theories. With this in mind and now that we considered some prominent FBAs, some objections against them, and potential replies to these objections, let us do likewise to TBAs.

Thought-based Approaches

Broadly, TBA theorists claim that emotions are necessarily constituted by thoughts, beliefs, judgments, and/or perceptions. Neu is one such TBA theorist. He starts by arguing to the effect that we cannot always distinguish emotions based on their expressions (Neu 2000, 14). Crying (though commonly associated with grief) is not the key distinguishing feature of grief. We can after all cry tears of joy. A mother's sacrifice for her child can also move us to tears. This is distinct from crying as a result of sadness or joy. The point here is that expressions and/or physiological changes are not always sufficient to allow us to distinguish between emotions. Neu argues that "... understanding the bodily expression of emotions will be more complex than noting a pattern of one physiological state triggering another" (Neu 2000, 16). It is our thoughts that allow us to decipher what emotions the expressions and/or physiological changes are associated with: "[t]he differences lie in the thoughts that provoke them or that, however inadequately, they express" (Neu 2000, 14). Given that it is our thoughts that allow us to distinguish one emotion from another (and not our expressions/physiological changes), this opens the way for Neu to argue that such thoughts constitute our emotions. He argues that "[t]houghts are crucial... in distinguishing one type of emotion from another. *They make each distinctively what it is.* Regret, remorse, shame, embarrassment, and a dozen other related states may all feel the same. What distinguishes each is the precise belief about what has gone wrong, about whether we are morally or in some other way responsible for it, whether we think others think less of us, and so on [emphasis added]" (Neu 2000, 12). Notice that the thoughts involved appear not to be yet-to-be-articulated impressions of an object. Rather, they appear to be propositional. It is the propositional beliefs that underlie the emotions that allow us to distinguish one emotion from another. As Neu argues, "[t]o say that thoughts are essential is to say, for example, that what is most distinctive about my anger is the belief (roughly) that someone has caused me harm (a belief presupposed by Aristotle's notion of a desire to 'return' pain) and that without that belief my state (no matter what my sensations) could not be one of 'anger' (after all, even if my stomach is churning in a typically angry fashion, that may be due to what I ate for lunch)" (Neu 2000, 12). If anger and fear have the same bodily

profile, then the propositional belief that ‘the actions of x caused me harm’ allows us to identify the associated bodily profile with anger. Unlike FBAs, Neu’s TBA does not face the distinction problem discussed above.

Another TBA is Nussbaum’s. She rejects James’ FBA as “too simple” as it “severs emotion from what is not only a necessary condition of itself [i.e. thoughts], but also a part of its very identity” (Nussbaum 2001, 30). The emphasis here is that, for Nussbaum, thoughts are necessarily constitutive of emotions. She sums it up by arguing that “emotions *always* involve thought of an object combined with thought of the object’s salience or importance [emphasis added]” (Nussbaum 2001, 23). Let us now consider what these constituent thoughts of emotions *involve*. For Nussbaum, “these emotions embody not simply ways of seeing an object, but beliefs – often very complex – about the object... In order to have fear... I must believe that bad events are impending... In order to have anger, I must have an even more complex set of beliefs: that some damage has occurred to me or to something or someone close to me; that the damage is not trivial but significant; that it was done by someone, probably, that it was done willingly” (Nussbaum 2001, 28-29). As we can see, the constituent thoughts of emotions involve perceptions and (on top of these) beliefs.⁴

The constituent thoughts of emotions also necessarily involve judgments. Nussbaum argues that emotions “involve judgments about important things, judgments in which, appraising an external object as salient for our own well-being...” (Nussbaum 2001, 19). Such judgments involve broadly two stages. *Stage 1* is the appearance stage. At this stage, when I receive news about the death of a loved one (for example), her death, her importance to me while she lived, her sudden absence, her being ‘extinct’ to my world appears as the objects of my grief. Such an appearance, Nussbaum argues, has “propositional content or at least combination: it combines the thought of importance

⁴ Nussbaum identifies a problem with associating emotions with beliefs. One may object that beliefs cannot be constitutive of an emotion. This is because of the possibility that an emotion may linger on despite my rejection of its underlying belief. Nussbaum puts it this way: “Jack may decide that he was wrong to accept his parents’ belief that African-Americans are ruining the country; but he still has intense anger against them” (Nussbaum 2001, 35). Nussbaum responds to this objection by arguing that one may still be gripped by false and contradictory beliefs: “Jack may be teaching himself some moral truths, but his deep-seated habits wipe those off the slate at times, and he is again in the grip of his past” (Nussbaum 2001, 36). I will return to this point later in this chapter.

with the thought of loss, its content is that this importance is loss" (Nussbaum 2001, 39). *Stage 2* is the assent stage. At this stage, if I reject the appearance of the death of my loved one or reject that she was important to me, there is no judgment. But if I accept "the way things look, take it into me as the way things are... the appearance has become my judgment, and that act of acceptance is what judging is" (Nussbaum 2001, 37).

Like Neu, Nussbaum argues that "it seems necessary to put the thought into the definition of emotion itself", if not "we seem to have no good way of making the requisite discriminations among emotion types" (Nussbaum 2001, 30). Introducing thoughts (involving judgments and beliefs) into what is constitutive of emotions, Nussbaum does not face the distinction problem. As with the above, we can see that the thoughts, judgments and beliefs can easily allow us to distinguish whether one is experiencing sadness or another emotion. Thoughts of loss (for example) are not primary when I am angry or disgusted. Such thoughts are primary when I experience sadness. This advantage of circumventing the distinction problem is common among TBAs. Robert Roberts (for example) uses the defining proposition methodology to individuate emotions. We can distinguish one emotion from another by examining their defining propositions. Here are some of them:

Anger: 'S has culpably offended in the important matter of X (action or omission) and is bad (is to some extent an enemy of what is good); I am in a moral position to condemn; S deserves (ought) to be hurt for X; may S be hurt for X' (Roberts 2003, 204).

Happiness: 'It is important that X be in condition Y; and X is in condition Y' (Roberts 2003, 279).

Fear: 'X presents an aversive possibility of a significant degree of probability; may X or its aversive consequences be avoided' (Roberts 2003, 195).

Sadness: 'X, whose continued existence or presence or availability is important, is no longer existent, present, or available' (Roberts 2003, 234).

As we can see, Roberts' defining proposition methodology appears to be an elegant tool for individuating emotions. The Schacter-Singer experiment poses no problem to his TBA. Anger is a response to an offence and happiness a response to an important expectation being met. It does not matter whether anger and happiness have the same bodily profiles. The methodology also

seems to make some FBA methods of individuation (that individuate emotions based on the associated feelings alone) look blunt. FBAs may find it difficult to distinguish between anger and resentment. But Roberts' TBA would encounter no such difficulty in identifying the subtle distinctions between these emotions. Compare the defining proposition for anger with:

Resentment: *'S has culpably offended in the important matter of X (action or omission) that touches me quite personally and is bad (is to some extent my enemy); I am in a moral position to condemn; I wish I could hurt S for X, but I am impeded from doing so directly'* (Roberts 2003, 214).

Both anger and resentment involve a culpable offence on the part of the object, a moral position to condemn on the part of the subject, and a recognition by the subject that the object deserves to be hurt. What distinguishes anger from resentment is the personal touch involved in the latter. The subject experiencing resentment wants to hurt the object personally. But the subject experiencing anger need not have such thoughts.

Returning to the TBAs of Neu and Nussbaum, we can also see that they do not face the intentionality problem (as faced by many FBAs). For Neu and Nussbaum, the experience of sadness is about something. To be sad involves thinking/judging/believing that I have lost someone or something important to me. My sadness is directed at the loss that I suffered. Unlike Damasio, my sadness is not directed at my internal bodily states. And unlike Prinz, my sadness does not represent loss (in a third-person, non-phenomenological sense), the emotion is actually directed at my loss (in a first-person, phenomenological sense).

Solomon's approach also makes clear the intentionality aspect of TBAs. He argues that emotions are *"about the world"* (Solomon 2004, 136). But he goes beyond this. He continues by arguing that '[e]motions are not just *about* (or *"directed to"*) the world but actively entangled in it... *emotions are subjective engagements in the world'* (Solomon 2004, 136). He does so to emphasize the point that the evaluative judgment necessarily constitutive of an emotion *"is not a detached intellectual act but a way of cognitively grappling with the world"* (Solomon 2004, 136). I draw the reader to this aspect of Solomon's TBA first to highlight his first-person, phenomenological aspect of intentionality. Given the robust sense of intentionality involved, perhaps we can see why some

would think FBAs are inadequate in this respect. It also highlights why Solomon is critical of TBAs like Neu's and Nussbaum's.

For Neu and Nussbaum, beliefs are constitutive of emotions.⁵ But Solomon argues that "belief isn't the right sort of psychological entity to *constitute* emotion. Beliefs are necessarily dispositions, but an emotion is, at least in part, an *experience*. Belief may be perfectly appropriate in *explaining* emotion but it is inappropriate in the *analysis* of emotion" (Solomon 2004, 140-141). The point here appears to be that beliefs are too dispositional to capture the essence of an emotional experience. Apart from beliefs, for Neu and Nussbaum, thoughts are also constitutive of emotions. Solomon also rejects this. He argues that "the notion of a 'thought' is too specific and involves too much intellect to provide a general account of the emotions... thoughts are too episodic for emotions, which often turn out to be enduring processes rather than mere episodes" (Solomon 2004, 141-142). There are two points to note here. First, thoughts are too cerebral to capture the characteristics of many emotions. Second, thoughts are episodic but emotions at times are processes (e.g. grief) and hence the former cannot accurately capture the essence of the latter.⁶ I highlight these criticisms to segue into whether/how the above TBAs can respond to these objections *and* how other TBAs can address these objections.

Let us start with Neu's TBA. Can beliefs (despite being dispositional) be constitutive of our emotional experiences? For Neu, we can sum up his answer with a rhetorical *why not?* I will use jealousy to elucidate what might be Neu's response. Anyone who has felt jealousy before and thought about the emotion can notice that apart from being episodic, it can be enduring as well. On my way back home, I may encounter my lover with a handsome man, a stranger. At this point a pang of jealousy may hit me (episodic jealousy). I hide from her, not wanting to see her, wanting to see how it plays out. Nothing happens and I continue to be on my way home. While on my way home, I mull on what had just happened. Thoughts like 'why didn't she tell me about him' or 'how

⁵ Solomon's criticisms in this paragraph targets Neu. However, his criticism is largely applicable to Nussbaum as well.

⁶ Solomon's criticisms of Neu and Nussbaum does not put him outside the category of TBA in general. Though (e.g.) Solomon rejects thoughts as constitutive of emotions, Solomon has a narrower conception of thoughts than what I envision in my thesis. I use the term 'thought' broadly to encompass not only beliefs, or ideas that pop up in our minds, but also judgments, perceptions, etc.

long has she known this man' comes to mind. Thoughts of losing her to him also pop up. This may continue for days as I keep my thoughts from her. I might feel angry at her too (as she should have told me about this). The more she keeps her silence about this man, the more I feel jealous, angry even. The point here is that jealousy can also be an enduring emotion. In light of this, let us consider Neu's argument on what is jealousy. He argues that jealousy is "compounded of fear and anger and is, at least in its erotic forms, importantly tied to love. In understanding jealousy, I would emphasize the constitutive role of thoughts... in particular thoughts involving fear of loss; more specifically, fear of alienation of affections (that is, loss of those affections to a third party, a rival), and, more generally and more deeply, fear of annihilation" (Neu 2004, 276). It appears that such jealous thoughts also involve beliefs. My jealousy may involve the belief that my lover is cheating on me. It may also involve the belief that I am inferior in some ways (perhaps in terms of good looks) to that handsome man. For Neu, though such beliefs may be dispositional, they can still be constitutive of such enduring emotional experiences. It seems that for Neu the dispositional nature of beliefs can still capture what an emotional experience is all about. We need not be conscious of the beliefs that underlie our emotional experience for us to experience it.

Now Neu also faces Solomon's objection that thoughts are too cerebral to be constitutive of all emotions. As discussed above, Neu seems to argue that *propositional* thoughts are constitutive of emotions. However, for Neu the propositional thoughts involved need not be conscious thoughts. Neu argues that:

"The thoughts involved in... emotions need not be conscious... Many thoughts that we ascribe both to ourselves and to others are less than fully explicit without being withheld from consciousness by dynamic forces of repression. Thus, I might explain to the police officer who stops me for jumping a light that 'I thought the light had turned green.' In such a case... I am not claiming that I explicitly thought 'the light has turned green now I can go forward,' any more than when I change gears I (as an experience driver) have to think explicitly 'I am in neutral and must now shift into first'" (Neu 2004, 299).

As such, Neu's response to Solomon may be that thoughts (constitutive of emotions) need not be conscious. And if so, Neu's TBA may resist the claim that his account of emotions is too cerebral. Though animals and infants do not have linguistically formulable thoughts, given that thoughts need not be explicitly propositional, Neu's TBA appears to accommodate some animal and infant

emotions.⁷ Neu's TBA also appears to offer an adequate response to Solomon's other objection that thoughts are too episodic for enduring emotional experiences. Neu argues that "[e]motions in general should not be confused with fleeting sensations. Emotions, like beliefs, are typically dispositional states that occur over time" (Neu 2000, 296). We can apply this point in response to Solomon's objection. It can be argued on Neu's behalf that emotions though constituted by thoughts should not be confused with the fleeting nature of sensations. Occurrent thoughts may come and go, but there may be dispositional thoughts as well (just as there are dispositional beliefs). When I am in an emotional state of jealousy, I may have an occurrent thought that my lover is cheating on me. Or I may be disposed to think she is cheating on me. This disposition can be triggered by (e.g.) objects that remind me of her. And if there are dispositional thoughts, then it can be plausibly said on behalf of Neu and in reply to Solomon that the former's TBA is not too episodic for enduring emotional experiences.

How then would Nussbaum's TBA fare against Solomon's objections? In response to the objection that beliefs cannot be constitutive of our emotional experience as they are too dispositional, Nussbaum can refer to what she describes as "ongoing or background emotions" (Nussbaum 2001, 70). She argues while referring to the fear of death that "one's judgments about one's own mortality and the badness of death persist... throughout one's adult life, though only certain circumstances bring them to consciousness; one's beliefs about the importance of one's own bodily health, and the vulnerability of one's body to disease and injury, similarly persist, unnoticed unless a circumstance calls them into view." (Nussbaum 2001, 70). For Nussbaum (like Neu), though beliefs may be dispositional, they can still be constitutive of such enduring emotional experiences. Since there is no need for us to be conscious of such beliefs when we experience the emotion, the dispositional nature of beliefs can still capture what an emotional experience is all about.

I also do not foresee Nussbaum having a problem with Solomon's objection that thoughts are too cerebral. She argues that "[s]ome emotions, even in an adult, may preserve a preverbal infant's archaic and indistinct view of the object. We therefore cannot think of all emotions as having a linguistically formulable content" (Nussbaum 2001, 79). If Solomon worries about thoughts being

⁷ This being said if for Neu emotions are necessarily constituted by beliefs, his TBA may still face the task of reconciling this with the apparent unlikelihood of animals and infants having beliefs.

too cerebral to provide a general account of emotions, Nussbaum eases this worry by opening the door to non-propositional thoughts being constitutive of emotions. As she argues “we should not understand the cognitive content of emotions to involve, in every case, anything like the acceptance of a linguistically formulable proposition. Many emotions, both nonhuman and human, involve only an evaluatively laden sort of seeing- as, where a creature sees an object as salient for its well- being. Where humans are concerned, such simpler emotions are particularly common in prelinguistic infants, but they can persist in adulthood as well, as many infantile emotions do” (Nussbaum 2016, 253). In terms of the objection that thoughts are too episodic to accommodate process-like emotions (hence they cannot be constitutive of emotions), I draw the reader first to Nussbaum’s distinction between “situational” emotions (where they are “fixed on a particular set of circumstances”) and “background” emotions (where they are “ongoing in the fabric of life (for example a fear of death that most people carry around with them), but can also become more concretely focused on a particular event (a particular threat to the person’s life)”) (Nussbaum 2016, 252). It appears that an emotional response elicited by an event can be understood as a situational emotion. But it appears that process-like emotions do not fit squarely within the confines of background emotions. This is because the former does not always remain in the background. Grief as a process-like emotion can remain in the background, but it can also manifest itself with sporadic bursts of tears triggered by encounters with objects that remind us of our loss. It appears then that process-like emotions are an amalgamation of situational and background emotions. With this in mind, we see that the episodic nature of thoughts can easily accommodate the situational aspect of process-like emotions. But it remains to be seen for Nussbaum how such thoughts can accommodate more enduring background aspects of process-like emotions.⁸ One may on Nussbaum’s behalf address this by distinguishing between occurrent and dispositional thoughts where the former (being episodic) can accommodate the situational aspect of an emotion and the latter its background aspect. For example, the explicit conscious thought of the loss of a family member can accommodate the situational aspect of grief. And the disposition to think of the loss of this family member when triggered by certain events can accommodate the background aspect of grief.

⁸ However, I note that this episodic problem is also faced by many FBAs where the understanding of emotions as feelings about our physiological changes are too episodic to adequately account for process-like emotions like grief. We may weep from time to time when we grieve, but we may still be in grief even if we do not weep or experience any suffering.

It is important again to note that I am not here to determine the victors of the above debates. What I intend to emphasize at this stage is that there has been a cause for concern in some aspects of the TBAs proposed by Neu and Nussbaum. Though they appear to adequately (or to some extent) address some concerns, there are other TBAs (like Solomon's) and FBAs that remain sceptical of Neu's and Nussbaum's approaches. I would like to pick up on three strands of such concerns to segue into a discussion about other TBAs. The first is the animal/infant emotion problem, i.e. if emotions are constituted by thoughts, such an account may exclude the emotion-like experiences of animals and infants. The second is the feelings problem, i.e. if emotions are not constituted by feelings, such an account ignores a paradigmatic feature of emotions. The third is the recalcitrance problem, i.e. if emotions are constituted by beliefs, would not recalcitrant emotions (where there appears to be conflicting beliefs) undermine such an account. Let us now start with the first problem. Roberts argues that "[h]ardly anyone will deny that dogs, squirrels, horses, chimpanzees, and many other species of animals experience at least some of fear, anger, jealousy, joy, hope (eager expectation), and grief. And human children well below the age of understanding or speaking a language can experience at least some of these emotions" (Roberts 2013, 89). And an emotion theory that does not accommodate animal/infant emotions seems impoverished in this respect. Notice that this is not a problem for the FBAs of James, Damasio and Prinz. By claiming that emotions are essentially constituted by feelings, they avoid the objection that their accounts of emotions are too cerebral. As such, FBAs like these can easily accommodate the emotions of animals and infants. Prinz, for example, argues that emotions *represent* organism-environment relations. This is a third-person account of intentionality (as opposed to a first-person phenomenological account). His FBA does not require us to provide any such first-person account of intentionality (that may be lacking in infants and animals). To be in fear (for example) there is no need for a thought/belief that *x* is dangerous. Now how would TBAs respond to the animal/infant emotion problem?

We have seen earlier in this chapter how Nussbaum would respond to the animal/infant emotion problem. I would now elaborate further. She starts off from the position that emotions are constitutive of thoughts that are appraisals. She then argues that though such thoughts are constitutive of emotions they need not be linguistically formulable. Infants can have such primitive thoughts and, as such, can have emotions as defined by Nussbaum. As she argues, "very young human infants, not yet capable of language, are still capable of many emotions, because they have

an inchoate sense of their own good and ill, and of the way in which objects and events contribute to that good or ill" (Nussbaum 2016, 251-252). As for animals, Nussbaum argues that "[m]ost animals make at least some appraisals of objects, from the point of view of their sense of their well-being, and have emotions in consequence. All that is required is that they see the object (a bit of food, say) as good from the point of view of the creature's own pursuits and goals" (Nussbaum 2016, 251). The thoughts that constitute emotions can (but need not be) propositional. Nussbaum's TBA, as such, can seemingly accommodate many infant/animal emotions.

One may find Nussbaum's response to be an ad hoc move. One may also find that her response might not be satisfactory in the sense that Nussbaum must clarify what she means as 'non-propositional thought'. This allows me to segue into Deigh's distinction between two general conceptions of thought. One conception is that thoughts necessarily involve propositional content. The other is that thoughts need not have propositions as their content, i.e. a mental state "in which the subject is cognizant of some object" (Deigh 2008, 43). Nussbaum's non-propositional thoughts seem to fall into the latter category. One question at this point would be *how can we better understand such non-propositional thoughts*. She does not appear to offer much help here, but one might say that non-propositional thoughts are 'unencoded' thoughts, i.e. complete thoughts yet to be expressed in a particular language. But Deigh argues that this will not work as logicians would refer to such 'unencoded' thoughts as propositions. As he argues, "[t]he concept of an encoded thought is that of a thought expressed in words of some language or its equivalent. When the thought is a complete one, then it is expressed by a complete, declarative sentence of that language. Consequently, if there is a version of this thought that is unencoded, it must be a complete thought in abstraction from every complete, declarative sentence that expresses it, and this is what logicians mean by a proposition" (Deigh 2008, 19-20). Instead he argues that we can find an answer in Locke's philosophy of mind, whereby the "thoughts we express in language are independent of our knowledge of language" (Deigh 2008, 22). We receive sensory input, compare them and abstract from them what interests us without utilising our linguistic capacities. If the Lockean conception of thought is correct, then perhaps TBAs like Nussbaum's can accommodate animal/infant emotion by incorporating such a conception of thought. As Deigh argues, Locke's conception "if it were sound, would close the gap between the way we typically understand emotions as intentional states and the way we typically understand primitive emotions" (Deigh 2008, 22). Deigh however raises a problem with this Lockean conception. He argues that it "fails

to account for the thoughts we express in language. Specifically, the power of abstraction, when understood as a power that operates independently of linguistic capacities, cannot yield such thoughts" (Deigh 2008, 22). The thought 'a coyote is not a dog' cannot result from just the powers of comparison and abstraction. Such a thought requires an understanding of certain linguistically mediated concepts like 'genotype'. If Lockean non-propositional thought is not a satisfactory account of non-propositional thought to some, Alex Grzankowski offers an alternative by arguing for "Non-propositional Intentionalism" to account for mental states like "loving, hating, and fearing" (Grzankowski 2016, 324). He starts by "differentiating states with propositional contents from states that are merely about propositions" (Grzankowski 2016, 316). As we can see, both states involve propositions. Grzankowski distinguishes them by arguing that the former has the proposition as "the *content* of the attitude" and the latter as "the *object* of the attitude" (Grzankowski 2016, 317). He then argues that "[w]hen an attitude has a propositional content, the attitude is sensitive to the truth of the proposition. The sensitivity varies across the attitudes: when the proposition in question is true, belief is *accurate*, desires are *satisfied*, perception is *veridical*... For instance, when one believes that p [that is by having p as the attitude's content], if p were true, things would be as one believes them to be. If one fears that p, if p were true, things would be as one fears them to be" (Grzankowski 2016, 318). Such propositional attitudes can then be contrasted with non-propositional attitudes. As Grzankowski argues:

Non-propositional attitudes "do not appear to have conditions of accuracy, satisfaction, and so on. Even when non-propositional attitudes have propositions as objects the states are not relevantly sensitive to the truth of the proposition. Suppose, for example, that John loves Sally. There seems to be no sensible question of the form, 'When is John's love accurate/satisfied/true?' Similar considerations apply to the other non-propositional attitudes. It is false that if one fears the proposition that p, then if p were true, things would be as one fears them to be. No notion of accuracy, satisfaction, or truth seems to get a grip in the cases of non-propositional liking, fearing, loving, hating, etc." (Grzankowski 2016, 318).

If Grzankowski's argument is correct, it appears we are progressing in the right direction in answering the question *how can we better understand non-propositional thoughts of emotions*. He does so partially by contrasting non-propositional attitudes with the features of propositional attitudes. Furthermore, Grzankowski's argument (if correct) would appear to allow for certain emotions to exist in a spectrum. For example, we can understand the fear of losing a job as a propositional

attitude (as it has propositional content, i.e. 'I fear that I would lose my job'). And we can understand a knee-jerk fear response to snakes as a non-propositional attitude (as it does not have a propositional content, i.e. 'Ah! Snake! Bad! Escape!') Given its lack of propositional structure, the latter unlike the former does not appear to have any accuracy/truth conditions. If emotions can exist in a spectrum, then perhaps the lower end of this spectrum can accommodate animal/infant emotions.

Now let us consider another TBA, that of Roberts. He argues that emotion is a kind of perception, a "concern-based construal". His theory would come under the category of perceptualist TBAs. Roberts describes such a construal in the following way: "You come into a situation that has emotional potential for you with a dispositional (or possibly occurrent) concern or desire, or an attachment; you then construe the situation in the terms characteristic of some emotion type, and the situation emotionally appears to you as it does because the terms in which you see the situation impinge on, connect with, that concern" (Roberts 2013, 46). We have also seen above how each emotion type has a defining proposition. To feel sad then is to construe the death of a loved one, one who was important to me, as my loss. Though Roberts' TBA does well in discriminating between emotions, it faces the animal/infant emotion problem. Tappolet argues essentially that concern-based construals are too cerebral to provide an account of emotions that covers animal/infant emotions (Tappolet 2010). Roberts in response argues first that "emotions are perceptions, not thoughts..." (Roberts 2013, 90).⁹ For him, there is no need for any conceptual nor propositional acquaintance for one to have emotions. Roberts argues that a dog can experience jealousy: "she doesn't have perception-independent *thoughts* corresponding to the roles [of the beloved and rival]. She's probably not capable of *reflecting* on her relationship and on the idea of its being alienated from her. But she's capable of making the needed distinctions among the roles and identifying the relational issue that is raised by this [rival's] snuggling of [her beloved]. She doesn't have to be a language-user to construe the situation in the way characteristic of jealousy" (Roberts 2013, 90). Furthermore, for Roberts, though emotions as mental states have a propositional structure, the person experiencing the emotion need not have in mind "explicit propositions" (Roberts 2009, 572). Emotions can be propositional in the sense that "they are

⁹ I continue to categorize Roberts' theory (and other perceptualist theories) as a TBA not because he argues that emotions are constituted by thoughts. Rather, I do so because concern-based construals or perceptions in general are closer to thoughts conceptually when compared to feelings. 'Thought' in TBAs is thus construed in a broad manner.

susceptible of reasonably accurate propositional characterization" (Roberts 2009, 573). In other words, emotions need only be indirectly (or from a third-person perspective) propositional. While experiencing an emotion, the person need not have in mind any proposition. What Roberts' theory requires is the possibility of a propositional form being fleshed out if/when the experience is subsequently explained by the person herself or an observer. Roberts argues that "this is the sense in which [he] claim[s] that the vast majority of emotions are propositional, whether their subjects are cultured adult human beings, six-month-old infants, or giraffes or chipmunks" (Roberts 2009, 573). It appears that Roberts has adequately addressed the animal/infant emotion problem. This being said, one may still be dissatisfied with Roberts' response. One reason for this dissatisfaction may be due to the subtleties behind the defining propositions of each emotion. For example, if the defining proposition of fear involves only the recognition that an object is something that I should avoid, we can reasonably infer from certain animal/infant behaviour that the animal/infant is experiencing fear. However, the defining proposition is far more refined. As we can see with the above, for Roberts, fear's defining proposition is: 'X presents an aversive possibility of a significant degree of probability; may X or its aversive consequences be avoided'. There is a question as to how we can reasonably infer that an animal/infant would experience a mental state that can be reasonably characterized this way. X may be perceived as an aversive possibility by a dog. But it remains open as to whether we can say that a dog perceived X as an aversive possibility of *a significant degree of probability*. Perhaps on Roberts' behalf one way of resolving this is to understand 'probability' in terms of 'likelihood'. Animals and infants may not have the language of probability, but they can perhaps experience a mental state along the following lines: 'X presents an aversive possibility of a significant degree of *likelihood*'. Perhaps we can infer this mental state by comparing how animals or infants respond to dangerous objects that are close and dangerous objects that are far away.

There are also perceptualist TBA theorists whose views are like Roberts'. Specifically, Julien Deonna, Tappolet and Fabrice Teroni argue that emotions are better understood as perceptions or analogous to perceptions (Deonna, Tappolet and Teroni 2015, 196). They find TBAs postulating that emotions are constituted by thoughts, judgments or beliefs to be defective. Rather, "in the same way as an object may visually strike you as being scarlet, it may emotionally strike you as being offensive or dangerous, and that the underlying similarities run sufficiently deep so as to warrant talk of emotion in terms of perception" (Deonna, Tappolet and Teroni 2015, 196). Let us

consider some details of such perceptualist TBAs. Tappolet argues that the key difference between such perceptualist TBAs and TBAs like Nussbaum's/Neu's is that "the former takes the representational content of emotions to be non-conceptual" (Tappolet 2016, 16). As such, perceptualist TBAs do not require an emotion to have propositional content nor involve evaluative concepts. Tappolet's distinction between the visual experience and judgment that a mountain range is jagged and blue details this point. She argues that:

"The visual experience and the judgment are both about the same mountain range, but intuitively, they represent their object and its properties in quite different ways. The visual experience is like a picture of the mountain range while the judgment is like a description involving terms that ascribe properties to the mountain range. In contrast with the case of judgment, it does not appear required to possess the concepts *jagged* and *blue* in order to have a visual experience of the jagged and blue mountain range" (Tappolet 2016, 17).

Now if emotions are perceptions so described, it appears that such perceptualist TBAs can easily accommodate animal/infant emotions. This is exactly what Tappolet argues: "Since the presentation of values [e.g. the fearsome] are non-conceptual, there is no difficulty in attributing emotions to non-human animals and young children (Tappolet 2016, 19). Animals and infants are unlikely to have a grasp of concepts, but Tappolet's TBA would not preclude them from having at least some emotional experiences.

I mentioned earlier that I would pick up on three strands of concern against TBAs. I addressed the first concern, the animal/infant emotion problem. Now I would address the second concern. TBAs appear to reject feelings as a necessary constitutive component of emotions. This rejection is problematic as it ignores a paradigmatic feature of emotions. As James argues above, there is nothing left of the emotions when stripped of their phenomenological content. It appears that we cannot genuinely say that we experience sadness (for example) without experiencing the feelings associated with sadness. Despite this, Nussbaum argues that "non-cognitive elements do not have the constancy and regular association with the emotion type in question that would be required if we were to include them in the definition of an emotion of a particular type" (Nussbaum 2016, 252). And as Neu argues, "James's approach, referring to physiology alone, leaves the possibility of discrimination [between emotions] unexplained" (Neu 2000, 18). Even if James' FBA fails to distinguish between emotions adequately, it appears that the price that Neu pays for a theory that

can individuate emotions is to reject feelings as a necessary constitutive component of emotions. This, as the feelings problem suggests, may come at too high a theoretical cost. A TBA theorist may explain anger in terms of its constitutive thoughts. When I am angry at someone, I may have the following thoughts: (a) I have “an unfavourable attitude towards people who make unjustified insults, (b) I believe that that someone insulted me without justification, and (c) I have “an unfavourable attitude towards [that someone] in so far as he has insulted [me] without justification” (Goldie 2000, 38). Peter Goldie argues that such an account of anger is consistent with me not experiencing any emotion at all (Goldie 2000, 39). I can have the belief that someone has insulted me. I can judge that someone has hurled an insult. I can have the desire to hit back at someone who has insulted me. And I can hit that someone. However, I can have all the above without feeling angry. We can tick all the boxes of what some TBAs would regard as constitutive of emotion (e.g. beliefs, thoughts, judgments), but yet not experience an emotion. Feelings appear to be a necessary constitutive component of emotions. In response to the feelings problem, Nussbaum argues that her provisional view is that “[t]here usually will be bodily sensations and changes of many sorts involved in grieving; but if we discovered that my blood pressure was quite low during this whole episode, or that my pulse rate never got above sixty, we would not... have the slightest reason to conclude that I was not really grieving” (Nussbaum 2001, 57). The point here first involves an implicit recognition that feelings are involved in paradigmatic cases of emotions. This may, for some, be a satisfactory response to the feelings problem as many, like Nussbaum, accept that in many cases feelings are involved. Instead, what Nussbaum rejects is that emotion concepts track feelings. This position seems plausible. We seem to be able to say that I am angry at that person who offended me even if ‘my blood is not boiling’. She argues that “[w]e may want to grant here that there are some nonintentional feelings that are frequently associated with a given emotion: take boiling and anger, or trembling and fear. Nonetheless, it appears that here too the plasticity and variability of people (both of the same person over time and across people) prevents us from plugging the feeling into the definition as an absolutely necessary element” (Nussbaum 2001, 60) If this is correct, then perhaps emotion concepts track something other than feelings. And if this is correct, then the feelings problem may not be that problematic. Feelings may not be essential to emotions after all.

However, Nussbaum’s response may be dissatisfactory to some on grounds that her explanation runs contrary to our phenomenological experience of emotions. Goldie argues that there are those

that postulate that “the emotion is taken as comprising feelingless belief and desire, plus something which is not directed towards the object of the emotion—a psychological add-on (a visceral feeling perhaps), or a purely physiological add-on (a visceral change perhaps), or both” (Goldie 2000, 40). It seems plausible that Nussbaum (at least her account of paradigmatic emotions) falls into this camp or a variant of this. She appears to be an ‘add on’ theorist, at least for paradigm cases of emotion. Goldie argues that add on theorists clearly got it wrong as “[a]cting out of emotion is not acting without emotion (explained by feelingless beliefs and desires) plus some added-on ingredient or ingredients. Rather, when an action is done out of an emotion, the whole action, and the whole experience of the action, is fundamentally different” (Goldie 2000, 40). In other words, to say that emotions are constituted by thoughts plus feelings is not, for Goldie, phenomenologically sound. If one accepts Goldie’s argument, then Nussbaum’s explanation is likely to remain unsatisfactory.

Now onward to the third strand of concern. This concern emerges out of the phenomena of recalcitrant emotions. An emotion is recalcitrant when it conflicts and persists against our better judgment. For example, I may come to believe (after considering the statistical probabilities) that using airplanes as a mode of transportation is not dangerous (the odds of crashing to my death is very low). However, I still stubbornly experience fear while flying. If emotions are constitutive of beliefs, then my belief that underlie my fear of flying is dangerous conflicts with my statistically based belief that flying is safe. The concern here would be that if there is such a conflict, would it not be that beliefs are not constitutive of our emotions? After all, we usually reject beliefs when we come to know that they are false. But with emotions, we seem to continue to hold on to such false beliefs. Nussbaum is aware of this recalcitrance problem and puts it this way: “It appears that people sometimes change their minds about the beliefs that underlie their emotions, but continue to have the emotions nonetheless... Does this mean, after all, that one may remove the evaluative beliefs without removing the emotions?” (Nussbaum 2001, 35). Her answer to this question is ‘no’. It does not follow from the phenomena of recalcitrant emotions that beliefs are not constitutive of our emotions. This is because for Nussbaum it is possible for us to have contradictory beliefs. And if this is correct, then the fact that the beliefs that underlie our emotions conflict with other beliefs does not undermine Nussbaum’s claim that emotions are constituted by beliefs.

Now Patricia Greenspan and Bennett Helm reject such a response. Helm argues that a TBA like Nussbaum's "assimilates the irrationality of the conflict between my knowledge and my fear to incoherence in judgment. This... does not make sense: conflicts between emotions and judgments do not verge on incoherence, for they are readily intelligible and happen all too often. What is needed... is a way of understanding the cognitive component of emotions without assimilating it to belief" (Helm 2001, 42). Notice how the recalcitrance problem has morphed. It has moved from the claim that *recalcitrant emotions undermine the view that beliefs are constitutive of emotions* to the claim that *the incoherence explanation of the recalcitrance of emotions is not a satisfactory one*. Greenspan offers a TBA response to this. She argues that though emotions may be constituted by thoughts, they are not constituted by thoughts *that are assented to*. Given that emotions comprise *un-assented* thoughts, we need not explain the conflict between emotion and judgment in terms of incoherence (Greenspan 1988, 3). Roberts also offers a somewhat similar solution to the recalcitrance problem. As seen earlier in this chapter, for him, emotions are construals. He argues that "it is better to think of emotions as a kind of appearance or *phantasia* or construal than as a kind of judgment, because emotions do have the character of appearance and often occur in the absence of assent to that appearance" (Roberts 2003, 89). Ruling out the element of assent, it seems that Roberts need not explain the conflict between emotion and judgment in terms of incoherence. After all, an emotion is only a *phantasia* and the emotion does not involve belief.

Helm argues that although Greenspan and Roberts avoid the use of the incoherence explanation, "it is not clear that they are thereby able to provide a proper understanding of the nature of the resulting irrationality" (Helm 2001, 42). There is some irrationality involved in a recalcitrant emotion: we can perceive *x* as dangerous (via fear) despite a contrary judgment that *x* is not dangerous. However, when we perceive a stick half submerged in water as bent despite the knowledge that it is not, we do not regard this as irrational. Roberts replies by rejecting Helm's objection. The view that emotions are concern-based construals can still explain the irrationality of recalcitrant emotions. They are irrational because "the knowing phobic feels torn between his judgment and his emotion in a way that the knowing stick-viewer does not feel torn between his judgment and his visual experience. The latter is complacent and normal, taking the illusion in stride; the former is in trouble and goes to a therapist... unlike the impression of the stick, the impression of the phobic object is a *concern*-based construal. The fear has a personal depth and life-disrupting motivational power that the illusion lacks" (Roberts 2003, 92). This is one way through

which perceptualist TBAs can respond to the recalcitrance problem. But this is not the only way. Sabine Döring takes a different route. Instead of showing how perceptualist TBAs can account for the irrationality that emerges from a conflict of emotion and judgment, she rejects the claim that recalcitrant emotions are irrational. There is as such nothing to account for. Essentially, Döring argues that “[j]ust like recalcitrant perceptions, recalcitrant emotions involve a cognitive conflict between experiential state and better judgement, and yet the subject is not irrational because he does not conflict himself” (Döring 2014, 124-125). Both the impression of the phobic object and the bent stick (when they clash with our better judgment) are not irrational. Helm thinks the former is irrational because he may be misled by how “recalcitrant emotions typically lead to practical conflict by motivating the subject to act in ways that interfere with the reasoned pursuit of his goals” (Döring 2014, 125). However, in both cases of phobic fear and the illusion of a bent stick, there is no irrationality. Both have cognitive conflicts and neither of them forces one to contradict oneself.

Grzankowski on behalf of TBAs who argue that “it is a necessary condition for being in an emotional state that one make a certain judgment or have a certain belief” offers another way of denying the irrationality of recalcitrant emotions (Grzankowski 2016, 641). He starts by arguing that the recalcitrance problem is a problem only when TBA theorists worry about attributing irrationality to a person experiencing recalcitrant emotions: “... if the subjects in question are not rational, the [recalcitrance problem] gets no footing. The ‘worry’ would amount to this: we must attribute conflicting beliefs to irrational subjects. This is no objection at all...” (Grzankowski 2016, 645). This would mean that the recalcitrance problem as described would not be a problem to Nussbaum who concedes that one who experiences recalcitrant emotions hold contradictory beliefs. But for TBA theorists who refuse this concession, Grzankowski argues that there are good grounds to question the recalcitrance problem described in the following manner: “[a]ny subject who believes that p and believes that not-p and does so by deploying the same (save for the negation) modes of presentation/concepts is not rational” (Grzankowski 2016, 646). Roughly, the argument here is that though a TBA defender may argue that “subjects who believe contradictory propositions may still be rational as long as they entertain those propositions using different concepts or by entertaining them under different modes of presentation”, an anti-TBA theorist can still insist that it remains plausible that “the recalcitrant subject is using the same modes across both fear and explicit belief [e.g.]...” (Grzankowski 2016, 645-646). Grzankowski offers two replies

on behalf of TBA defenders. However, for brevity's sake, I would focus only on one. He argues that "the prima facie reason for thinking the contents are entertained under the same modes is that it seems to subjects [via introspection]... that they are. But we know that things are not always as they seem, so [TBAs] are well within their dialectical rights to deny that [the subjects experiencing recalcitrant emotions are] using same modes of presentation or concepts" (Grzankowski 2016, 649). In short, TBA defenders are open to argue that until and unless critics show that those experiencing recalcitrant emotions believe contradictory propositions by using the same modes of presentation/concepts, they need not concede that such people are being irrational.

At this point, we discussed some prominent FBAs and TBAs along with some objections against them. Before we consider the third camp of emotion theories, i.e. HYAs, we can start by taking stock on what FBAs and TBAs do well and not so well with. This would serve as a precursor to drawing out the presuppositions that I would take into the subsequent chapters. We would then pay special attention to how HYAs attempt to avert some problems faced by FBAs and TBAs. Once we do this, we would be at a stage where the presuppositions are evident. Once we reach this stage, we need only clarify what these presuppositions are and explain why they are important for the chief aim of my thesis.

Hybrid Approaches

Now what do FBAs do well? They do not face the feelings problem (as emotions to them are essentially feelings). They do not face objections on how beliefs may be too dispositional and how thoughts may be too cerebral for an account of emotions (as neither beliefs nor thoughts are constitutive of emotions). Neither would FBAs face the animal/infant emotion problem (as emotions are essentially noncognitive). As for the recalcitrance problem, though we did not discuss this earlier with FBAs, there are two ways to go about this. First, if recalcitrant emotions are not irrational, then FBAs would not face the recalcitrance problem as there is no irrationality to explain. Second, if recalcitrant emotions are irrational, then FBAs would face the same task of explaining how a conflict between noncognitive emotions and judgments can be considered irrational. This segues us into what FBAs do not do so well. They face the meaningfulness problem. FBAs would have to explain how noncognitive emotions can be meaningful. They also face the

intentionality problem. FBAs have to take the extra step to show how emotions can be intentional. FBAs might not do so well with the distinction problem either. FBAs have to reconcile the extensive list of emotions with a limited set of physiological changes. At this point, I would segue into what TBAs do well. They do not face the meaningfulness problem (as emotions are constituted by thoughts/judgments/beliefs/perceptions that provide an easily accessible source that can explain the emotions' meaningfulness). For nearly the same reasons, TBAs do not face the intentionality problem either. Furthermore, given that (for TBAs) emotions are not essentially feelings of physiological changes and thoughts/judgments/beliefs/perceptions provide us with the tools necessary to distinguish between emotions, TBAs do not face the distinction problem. However, TBAs also have their own kryptonite. They might not do so well with the feelings problem. They have to respond to the question of how thoughts/judgments/beliefs/perceptions that lack affect can be constitutive of emotions, but feelings are not. Some TBAs may stumble at the objections that beliefs/thoughts are too dispositional/cerebral to be constitutive of emotions. They might also stumble at the animal/infant emotion problem (where emotions (as understood by TBA theorists) may be too cerebral to cover animal/infant emotions). Lastly, TBAs face the recalcitrance problem regardless of whether recalcitrant emotions are considered irrational.

Notice with the above that there is some symmetry between what FBAs do well and what TBAs do not (and vice versa). For example, FBAs do not encounter the feelings problem, but TBAs do. And TBAs do not encounter the intentionality problem, but FBAs do. The question that may arise from this is what if there is an emotion theory (apart from FBAs and TBAs) that can accommodate both feelings and intentionality? In providing answers to this question, we see the emergence of HYAs.¹⁰

A prominent theory that I would focus on is Goldie's HYA. Broadly, he is dissatisfied with both FBA and TBA accounts of emotion. Against TBAs (like Nussbaum's and Neu's), Goldie argues that beliefs and judgments by themselves are not sufficient to constitute an emotion. He also expresses dissatisfaction toward add-on solutions to this problem. Emotions are not beliefs +

¹⁰ I note that it is not my aim to show how successful HYAs are in addressing the problems. Rather, I aim to provide an overview of HYAs by showing how they can address such problems. By doing so, it is one way of allowing me to segue into making clear the presuppositions that I would carry into the subsequent chapters.

judgments + feelings. This is because, for Goldie (as seen above), such an account runs contrary to the phenomenological experience of emotions. Against FBAs (like James'), Goldie first argues that he is "in agreement with the *spirit* of what James says, and the emphasis he puts on feelings". But he then argues that James' FBA leaves "no obvious place for *feelings towards*... [emphasis added]" (Goldie 2000, 54). For Goldie, emotions are not just feelings of physiological changes. They are directed at objects.¹¹ Similar to Prinz, Goldie's HYA accommodates a more abstract understanding of the *object*. It is something that our thoughts and feelings are typically directed toward. It need not be a specific person or event. It can be the relationship between persons and events/things. For example, the object of one's pride is the relationship between oneself and the work accomplished. Though an emotion is an intentional state, we cannot equate it with other intentional states like beliefs or desires. If we equate it with such states, we risk leaving feelings out of emotions. To capture both the intentionality and feeling involved in emotions, Goldie introduces the concept of *feeling towards* (Goldie 2000, 19). For Goldie, feeling towards is "thinking of with feeling" (Goldie 2000, 58). He argues that "emotional feelings are inextricably intertwined with the world-directed aspect of emotions, so that an adequate account of an emotion's intentionality... will at the same time capture an important aspect of its phenomenology" (Goldie 2002, 242).¹² As we can see here, if Goldie's argument is correct, then we would have an HYA that averts the intentionality and feelings problems altogether.

What then about the animal/infant problem? How would Goldie handle this? Prima facie, the concept of feeling towards introduces a cognitive aspect to emotion. If to feel an emotion is to think with feeling, it appears this may be too cerebral a definition for the emotion-like experiences of animals and infants. To the contrary, Goldie argues that "there is a more primitive story to be told about many of our emotional responses, particularly those which, intuitively, we have in common with many other animals" (Goldie 2000, 46). He provides an example of how we would likely react in fear and throw ourselves away from the bus that is about to hit us. He argues that "the beliefs and desires which make sense of [such an] emotional response do not also causally explain the emotional response" (Goldie 2000, 46). The belief that a bus headed our way is a danger to my safety and the desire to avert danger may not have caused our fear (as they may have come too

¹¹ Hence, Goldie (at least implicitly) endorses what TBAs do well in, i.e. adequately addressing intentionality.

¹² Helm also argues for a similar position. He argues that emotions "are intentional feelings of import" (Helm 2009, 254). Given that his position would address the intentionality problem in a fashion similar to that of Goldie, I would not detail Helm's account.

late for the emotion). Instead, Goldie argues that the relationship between recognition of an object (as dangerous, e.g.) and the emotional response (e.g. fear) is “meant to be sufficiently protean that it can accommodate, at the one extreme, psychological episodes as complex and sophisticated as those involved in the fear of being made redundant, and, at the other extreme, psychological episodes as primitive and *animal-like* as those in fear of the oncoming bus” (Goldie 2000, 47). Goldie’s point is that even adult humans experience animal-like emotions. An emotion theory would be inadequate if it cannot accommodate such animal-like emotions. Goldie argues that his HYA can accommodate such animal-like emotions in humans (as the cognitive threshold for a psychological state to be considered as an emotion is not a high one). If he is correct, then Goldie’s HYA can also accommodate for many emotions in animals and infants. Perhaps it cannot be said then that Goldie’s HYA is too cerebral for animal/infant emotions.

As for the recalcitrance problem, Goldie argues that in such circumstances “[r]ather than thinking of [a person with recalcitrant emotions] as, simply, irrational, [he] find[s] it more helpful to bring to bear the notion of *cognitive impenetrability* in emotional experience...” (Goldie 2000, 76) It appears that his potential response to the recalcitrance problem is a practical one. The term *cognitive impenetrability* may be more apt to explain why we sometimes cannot but have a particular feeling toward *x* despite our judgment to the contrary. It is not that we have contradictory beliefs (in recalcitrant emotions we have *feelings towards* conflicting with a belief instead). We are not contradicting ourselves. Rather, it may be that “[we] have the right beliefs yet, try as [we] might, [we] are still not able to resist feeling [insulted, for example, at a non-insulting remark] and react accordingly” (Goldie 2000, 76). If this view is correct that Goldie’s HYA addresses the recalcitrance problem by rejecting that there is anything irrational to begin with.

As for the meaningfulness and distinction problems, given that *feeling towards* involves thinking, this provides us with the tools to explain the meaningfulness of and distinguish between emotions. Goldie goes beyond this to argue that emotion “is a complex state, relatively more enduring than an emotional episode, which itself includes various past episodes of emotional experience, as well as various sorts of disposition to think, feel, and act, all of which can dynamically interweave and interact. What holds these diverse elements together is their being part of a narrative” (Goldie 2000, 19). Notice how Goldie distinguishes an emotion from an emotional episode. An emotion is not an

emotional episode and the former can comprise a series of the latter. Goldie's HYA (in particular his understanding of emotion as a series of psychological states held together by a narrative) may provide us with an additional and seemingly useful tool to explain the meaningfulness of emotions and discriminate between different emotions.

Similar to Goldie's concept of *feeling towards*, Linda Zagzebski argues that "an emotion is a state of feeling a certain way *about* something or *at* something or *toward* something of a certain description" (Zagzebski 2004, 60).¹³ She continues to argue that "[t]he emotion bears an external relation to its intentional object and an internal relation to the conscious construal of the intentional object in the emotional state" (Zagzebski 2004, 60). We can consider her theory as an HYA as she proposes emotion as a cognitive/affective state: "Other philosophers have argued that emotions combine cognitive and affective aspects, but [her] position is that the two aspects cannot be pulled apart" (Zagzebski 2004, 59). An emotion is also a "unitary state" where one cannot feel the emotion of pity without seeing something as pitiable (Zagzebski 2004, 68). What is clear with the above is that for Zagzebski emotions are neither purely cognitive states (like beliefs, judgments) nor purely affective states (like feelings about physiological changes). This means that Zagzebski's HYA would not likely face the feelings problem (faced by some TBAs) and intentionality problem (faced by FBAs).

Now as we can see, an emotion for Zagzebski has cognitive content. Let us probe deeper into this. She argues that "[w]hen an agent is in the state of emotion, she sees the intentional object of the emotion as falling under... 'thick affective concepts.' Corresponding to each emotion there is a thick concept under which the agent represents the intentional object of her emotion.... Pity is an emotion consisting in feeling pity for something seen as pitiful... fear is an emotion consisting in feeling afraid of something seen as dangerous..." (Zagzebski 2004, 61). In light of this, Zagzebski's HYA seems to have readily accessible material that would assist in both explaining the meaningfulness of emotions and distinguishing between different emotions. Given that to fear is to see something as dangerous (as opposed to just feeling my rising heartrate), Zagzebski's HYA

¹³ The key difference between Goldie's HYA and Zagzebski's is that for the former emotion comprises not only feeling towards (instead it comprises a series of elements weaved together by a narrative). But for the latter, this state of feeling towards by itself is the emotion. For Goldie, an emotional episode is not an emotion. But for Zagzebski, it is.

portrays emotions as not just being meaningful in general, but also meaningful from the perspective of the person who experiences it. And given that each emotion picks up a unique thick affective property in its intentional object, it appears that we can use the thick affective property as a touchstone to individuate emotions. If I see something as dangerous, I am experiencing fear. If I see something as contemptible, I am experiencing contempt. And so on.

Zagzebski's HYA also distinguishes an emotional state from the psychological state of belief. She argues that "[w]hen I see something as dangerous, contemptible, charming, offensive, and so on, I am in the corresponding emotional state, but emotion does not necessarily include having a belief about the intentional object of the emotion" (Zagzebski 2004, 75). We may not be sufficiently conscious of our emotions to allow beliefs to take shape. And even if we are conscious of our emotions, we are still open to distrust it. The point that I would like to draw from the above is that Zagzebski's HYA escapes objections like 'beliefs are too dispositional to capture what is essential to an emotional experience'. Beliefs, for Zagzebski, are not necessarily constitutive of emotions. This segues us into a discussion on whether her HYA would be too cerebral to provide a general account of emotions. For Zagzebski, emotions "can be expressed propositionally" (Zagzebski 2004, 74). But it does not follow from this that emotions are propositional attitudes (like say beliefs). She is also clear that an emotion does not include a judgment or a belief. She argues that "[s]omething can appear fearsome to the agent when she does not judge that it is fearsome. In fact, she may judge that it is not fearsome" (Zagzebski 2017, 33). Zagzebski also argues in relation to the emotion of admiration that "we can recognize a good person or good personal behavior without having a prior concept of good against which a given instance is measured. In the simple cases, we... admire something that we then judge, defeasibly, to be good" (Zagzebski 2004, 54). The point that I would like to draw out is that admiration is (for Zagzebski) an emotion that picks out good people/behaviour. More importantly, admiration can pick out good people/behaviour without any prior conceptual understanding of what is good. If her view is correct, it appears that Zagzebski's conception of emotions is not too cerebral to provide a general account of emotions. There may not be a prior need to grasp concepts usually associated with an emotion to experience the emotion. If so, Zagzebski's HYA may also accommodate a wide range of animal and infant emotions.

In terms of the recalcitrance problem, I alluded earlier to Zagzebski's view that it is possible to feel fear toward x but yet judge that one has nothing to fear about x . On this point, she argues that in such a circumstance, "there is a misstep of some kind – a lack of fit between her faculties and her environment. The faculty or disposition through which something appears to her in her emotional state is misrepresenting the object" (Zagzebski 2010, 33). It is not clear whether Zagzebski thinks recalcitrant emotions are irrational. On one hand, this lack of fit appears to be able to be construed as a form of irrationality. On the other, given that emotions are not constituted by belief, we can say that there is no irrationality here as the person who experiences a recalcitrant emotion is not contradicting herself by having contradictory beliefs. I would not speculate further on this and would leave it open whether Zagzebski's HYA can address the recalcitrance problem.

Now with the above discussion, we have seen how some prominent HYAs aim to address the problems faced by TBAs and FBAs. Whether HYAs are successful in doing so is not a question that I would look into. The point of the above exercise is to highlight the key features of some HYAs by identifying the problems faced by TBAs and FBAs and showing how HYAs can address these problems.¹⁴ Also, I have not addressed objections to the above HYAs. But I do not think the aim of my thesis requires me to do so. As we would see in the subsequent and final section of this chapter, this chapter centres on drawing out certain problems faced by TBAs and FBAs and how TBAs, FBAs and HYAs can address these problems. I do this with the intention of drawing to the readers' attention the presuppositions that I would take into the subsequent chapters. As such, no discussion on objections to HYAs is required for this purpose (despite there being several objections). Now with this in mind, onward to the next section!

¹⁴ The reader would probably have noticed that we did not discuss several HYAs. For example, Aaron Ben-Ze'ev argues that one way of answering the question *what is an emotion* is to describe "all the basic, features, characteristics, and components" of the emotion. In terms of the components of an emotion, they are "cognition, evaluation, motivation, and feelings" (Ben-Ze'ev 2010, 60-61). His theory as such is an HYA in that it includes within the definition of a typical emotion elements that TBAs and FBAs would include. The exclusion of HYAs such as these are due to mainly two reasons. First, though such theories are important in their own right, they do not serve a practical purpose in my thesis as they would face similar problems to that of FBAs and TBAs (and we have already discussed the problems important to progress my thesis). Second, a discussion of such HYAs need not be present to draw out the presuppositions that I would take into the subsequent chapters.

Presuppositions

Now with prominent FBAs, TBAs and HYAs identified, objections against FBAs and TBAs raised, and responses from FBAs, TBAs and HYAs produced, I will discuss the presuppositions that we can draw out of these objections that I would take into my subsequent chapters. I do so by keeping in mind the aim of showing how EAMs can explain moral unanimity and diversity. These presuppositions are important as they not only exhibit my ecumenical approach toward emotion theories but are also aimed at enhancing the plausibility of EAMs and/or their ability to accommodate moral unanimity and diversity. Some presuppositions do so by excluding criticisms (grounded on a preference of one emotion theory over another) of EAMs as a plausible metaethical theory. By demoting to a presupposition the claim ‘emotions are not just brute feelings’ (e.g.), this move may preclude a proponent of James’ FBA from objecting that emotions (as feelings produced by bodily changes) cannot be necessarily constitutive of moral judgments. The presuppositions can also help present EAMs as a plausible approach toward explaining moral unanimity and diversity without taking a stance on which emotion theory is the correct one. By presupposing that some emotions have some cognitive element that can be influenced by thought (e.g.) can help in providing an EAM explanation on moral diversity. Without further ado, let us consider these presuppositions with more detail.

First, I presuppose that the emotion theories considered in my thesis can address adequately the meaningfulness problem. This would exclude certain EAM sceptics, i.e. those who question how meaningless emotions could constitute our moral judgments. What I wish to draw out with the meaningfulness problem is that I would take for granted that emotions cannot be *just* brute feelings. I would have to make clear here that by making this move I am not taking sides in the debates between cognitivists and non-cognitivists pertaining to emotions and moral judgments. As we have seen above, emotions can arguably be noncognitive but yet meaningful. What I intend to do is much more modest, i.e. to take for granted that emotions are more than just feelings. Second, I also presuppose that emotion theories can address the distinction problem adequately. This presupposition is important as the EAMs I consider in the next chapter assumes that emotions are sufficiently distinguishable. For example, the moral judgment that x is offensive may be grounded on anger (i.e. a sufficiently distinguishable emotion).

Third, the emotion theories that I have in mind can accommodate animal/infant emotions. This presupposition is important as I would rely on such emotions (minimally influenced by cultural practices, norms, etc.) to explain how EAMs can explain moral unanimity. More specifically, if some animals and infants panculturally share a similar set of emotions with similar eliciting conditions (prior to any or with very little cultural learning), this may provide us with some basis to provide an explanation of moral unanimity. Fourth, emotion theories should adequately capture the intentionality of emotions. I would rely on how fixed abstract objects associated with emotions (e.g. pity's object is the pitiable) can explain some level of moral unanimity. Roughly, if pity's object is always the pitiable and pity is constitutive of some moral judgments, this may be a step towards explaining near universal judgments like 'inflicting needless pain on others is wrong'. Understanding emotions as intentional would also allow me to explore how varying concrete objects associated with emotions (e.g. anger's abstract object is the offensive, but what is specifically offensive varies) can explain moral diversity. If anger is constitutive of some moral judgments and what is offensive varies and can vary from person to person, place to place, this may be a step towards explaining the diversity of judgments on what is offensive. A Korean (for example) may get angry at a younger person who did not bow before her and judge that this younger person has done something offensive. For others, non-bowing may not be an eliciting condition of anger and no judgment of offence would ensue.

Fifth, despite the objections against thoughts being constitutive of emotions, I would presuppose that some emotions are at least constituted by some cognitive element. I remain neutral however on what exactly this cognitive element is. I would not make a determination on whether these emotions are constituted by beliefs, judgments, perceptions, or evaluative feelings. Why then is this presupposition important? The short answer is that the presence of some cognitive element in some emotions is a step toward explaining moral diversity. The longer answer is that the presence of some cognitive element in some emotions opens up the possibility for other thoughts to influence the cognitive content of these emotions. For example, if belief p is constitutive of emotion e , then a subsequent discovery that $not\ p$ is true can result in one's discard of belief p and adoption of belief $not\ p$. With belief p discarded, one may no longer experience emotion e . As such, the judgment that may usually ensue from emotion e may no longer be there. At this point, I would have to note that the situation of discarding old and adopting new beliefs is far more complex than I have described. The belief that 'John has insulted me' can be challenged from multiple vantage

points. This belief may have been based on hearsay, and it might turn out that John never said those insulting remarks. Such a discovery would undermine my initial belief. There may also be other background beliefs that underlie my initial belief (for example, the belief that the words uttered by John (if uttered by John) is offensive). Such background beliefs are often culturally informed. If so, such background beliefs can be undermined by the culture that I am in (cultural beliefs on what is insulting might shift) or other cultures (other cultural beliefs that are at odds with my cultural belief). The point that I am trying to draw out is that if emotions comprise a cognitive element and this element can be shaped by cultural beliefs, this is a step toward explaining moral diversity.¹⁵

In sum, this chapter outlined the three main emotion theory camps: FBAs, TBAs, and HYAs. The objections against FBAs and TBAs (and how we address them) draw out the above presuppositions that I would take into the subsequent chapters.

¹⁵ The recalcitrance problem, however, draws out an aspect of the emotions that appears to pose as a problem to an EAM attempt to explain moral diversity. If some emotions are cognitively impenetrable (and these emotions ground our moral judgments) would this not be an obstacle in explaining the diversity of moral judgments? At this point it would suffice to note that though the Müller-Lyer illusion is often used by perceptualist TBAs as analogous to how emotions as perceptions might be illusory, the same perceptualist TBAs may concede that the analogy stops at the point where emotions though somewhat cognitively impenetrable are *not totally so*. This is unlike the Müller-Lyer illusion where I would think we could never perceive both lines being of the same length. My point here is that despite recalcitrant emotions, they appear to be able to be influenced by other thoughts and beliefs. As Goldie argues, “cognitive impenetrability admits of degrees, and need not be total...” and “cognitive therapy with regard to fear of flying has been shown markedly to reduce [phobic] fear; if people face up to what they are afraid of, and to the risks involved, then things go better for them when they fly” (Goldie 2000, 77).

Chapter Two

“All morality depends upon our sentiments; and when any action, or quality of the mind, pleases us after a certain manner, we say it is virtuous; and when the neglect, or non-performance of it, displeases us after a like manner, we say that we lie under an obligation to perform it. A change of the obligation supposes a change of the sentiment; and a creation of a new obligation supposes some new sentiment to arise”
(Hume 1896, 517).

EAMs

With emotion theories set out, I can now discuss EAMs. EAMs are united by the thesis that emotions are in one way or another essentially involved in moral judgments. To facilitate my aim of illustrating how EAMs can accommodate moral unanimity and diversity, the following are some tokens of how this ‘essential involvement’ can pan out:

- (1) Moral judgments (as utterances) express emotion.
- (2) Moral judgments (as propositions) have emotion as referents.
- (3) Moral judgments (as a psychological state) are constituted (at least in part) by emotion.¹⁶

The reader would notice that there are three meanings of ‘moral judgment’ involved. (1) and (2) refer to moral judgments as utterances/propositions. Whereas (3) refers to moral judgments as psychological states. It comes as hardly surprising that ‘moral judgment’ means different things to different people. As Garrett Cullity argues, “[t]he term ‘moral judgement’ can refer to an activity, a state, a state-content, a capacity or a virtue” (Cullity 2011). As we can see, we can understand moral judgments in at least five ways. For the purposes of my thesis, I would be referring to the first three. Moral judgment as an activity is the process of “thinking about whether something has a moral attribute”. The result of this process can be a psychological *state*, i.e. the “state of judging

¹⁶ An interesting implication of my approach is that it allows for the possibility of moral judgments (as utterances/propositions) to be expressions of moral judgments (as a psychological state that is at least in part constituted by an emotional state). As odd as it sounds, moral judgments can be expressions of moral judgments.

that the thing has the attribute". And the state is distinguishable from its *content*, i.e. "what is judged by me, rather than my judging it" (Cullity 2011). In light of this, we can understand types (1) and (2) above as state-contents (or representations of state-contents) and type (3) as states. By doing so, we exclude an array of moral sentimentalisms from the purview of EAMs strictly so called. This exclusion is done to present a well-defined subset of EAMs. Without this well-defined subset in mind, it appears that I would face an insurmountable task (from a practical standpoint) in arguing that EAMs *in its myriad types* can accommodate moral unanimity and diversity. The difficulty involved is likely to be exacerbated if I am to address some objections against all different types of EAMs and to some extent consider responses to them. Apart from intending to avoid a practical difficulty, the narrowing of the scope involved may allow me to adopt an ecumenical approach in several debates. For example, by excluding epistemological EAMs from the discussion, I can remain neutral on whether moral judgments are ultimately justified by emotional responses. Attending to such debates would likely be a distraction to the ultimate aim of my thesis.

Allow me to explain the motivation behind my approach. I am motivated to present 'moral judgments' in terms of *utterances/propositions* considering the role language plays in moral discourse. Language allows us to be explicit about our moral stances. If I am incensed by (e.g.) how female genital mutilation is permitted or even encouraged in some parts of the world, I can put my emotional state into words. If we are explicit about our moral stances, we can also defend them. I can give reasons to support my stance. Others can criticize my stance and vice versa. Others can also give reasons for the stance I reject. In short, if we see value in moral discourse, discussing 'moral judgments' in terms of utterances/propositions is useful. And if we see value in moral discourse, its value is likely to be enhanced or made more evident considering the myriad cultures that we encounter. Moral discourse and moral language allow people from myriad cultures to explicitly state their stances. Moral language can reveal moral unanimity (the stances they share) and moral diversity (the stances they do not share). In other words, speaking of 'moral judgments' in terms of utterances/propositions and their connection with emotion can facilitate an important aim of my thesis, i.e. to illustrate how EAMs can explain moral unanimity and diversity. As for presenting 'moral judgments' in terms of a psychological state, I am motivated to do so as it allows me to segue into the various empirical studies that largely see moral judgments as a psychological state. These studies allow me to probe into questions like what evolutionary mechanisms influence our moral judgments (as a psychological state) and how we can/have 'departed from' such an

influence. Both questions (as we shall see in chapters three and four) help in explaining instances of moral unanimity and diversity.

With the above in mind, I will start the next section by outlining some empirical support for EAMs. After which, I would present the following token EAM theories: emotivism, expressivism, empathy-based sentimentalism, dispositionalism and pluralism.

Some Empirical Support

I am motivated to discuss some empirical support for EAMs for mainly four reasons. First, the empirical claims that follow when taken individually and especially collectively help build the case for EAMs. Second, the discussion serves as a metaphorical separation of sheep from goats. Both the empirical and EAM claims highlight the nexus between emotion and moral judgment. But this is where their similarity stops. To claim that emotion co-occurs with moral judgments, for example, is distinct from claiming that emotions constitute moral judgments. Third, this metaphorical separation of sheep from goats can help clarify what token EAM theories are. It seems to me plausible that we can understand token EAM theories better by understanding what these tokens are not (especially if both the empirical and EAM claims aim to establish some nexus between emotion and moral judgment). Fourth, by enhancing our understanding of token EAM theories in the manner described, it can facilitate the ultimate aim of my thesis to show how EAMs can explain moral unanimity and diversity. Here are the empirical claims that I would discuss:

- Emotions *co-occur* with moral judgments;
- Emotions *influence* moral judgments;
- Emotions are *sufficient* for moral judgments; and
- Emotions are *developmentally necessary* for moral judgments.

Emotions co-occur with moral judgments. As Prinz argues, “[i]t is fairly obvious from experience that when we judge that a moral rule has been violated, we typically have a negative emotional response. This piece of introspective psychology has been confirmed again and again, in every study of what goes on in the brain during moral judgment” (Prinz 2006, 30). This claim supports

EAMs to an extent. It makes clear the correlation between emotions and moral judgments. This correlation can be a step toward establishing any token EAM theory. The following counterfactual would make this point clear. If emotions do not at all correlate with moral judgments, it appears difficult to see how we can establish (e.g.) the claim that moral judgments express emotions or have emotions as referents. This being said, the claim that moral judgment express emotions or have emotions as referents does not necessarily follow from the co-occurrence claim. Furthermore, we can easily flip the co-occurrence claim to support (e.g.) non-EAMs. Instead of moral judgments expressing an underlying emotional state, non-EAMs can argue that moral judgments *cause* emotional responses. Instead of anger at the murderer being the underlying psychological state of the moral judgment 'murder is wrong' (e.g.), the moral judgment 'murder is wrong' can be logically independent from anger in the sense that the former causes the latter. As such, we can see that though the co-occurrence claim and EAMs are united in establishing a nexus between emotion and moral judgment, the former by itself does not sufficiently offer support to EAMs. The co-occurrence claim 'needs help' from other empirical claims. In other words, the co-occurrence claim necessarily supports EAMs, but its support is insufficient.

Emotions influence moral judgments. Prinz argues that "[a] negative emotion can lead us to make a more negative moral appraisal than we would otherwise have" (Prinz 2006, 31). Again, this claim establishes a nexus between emotion and moral judgment. As argued earlier, a claim that establishes a correlation between emotion and moral judgment can support EAMs. But more than this, for Prinz, this influence claim supports EAMs in the sense that the influence claim reveals a phenomenon that can be explained by EAMs. As he argues:

"If the judgment that something is wrong contains indignation, then becoming indignant would promote that judgment. By analogy... imagine that when we judge something to be amusing we are making a judgment of the form 'that thing causes this state,' where 'this state' is an inner demonstrative pointing to amusement. Becoming amused promotes the judgment that something is amusing by furnishing us with one of its constituent parts" (Prinz 2007, 23).

I will now consider one out of many experiments that Prinz uses to support his claim that moral judgments comprise emotions. First, participants were hypnotized to feel disgust when they encounter an arbitrary word in vignettes. Thalia Wheatley and Jonathan Haidt found that the

participants would make more severe judgments when the word was present in the vignettes (Wheatley and Haidt 2005). For example, in this vignette: “Congressman Arnold Paxton frequently gives speeches condemning corruption and arguing for campaign finance reform. But he is just trying to cover up the fact that he himself [will take bribes from/is often bribed by] the tobacco lobby, and other special interests, to promote their legislation” (Wheatley and Haidt 2005, 781), participants who were hypnotized to feel a pang of disgust when they hear the word ‘take’ judged the Congressman’s action to be more disgusting and more morally wrong when the word was present. The emotion (triggered by the word and not the situation described in the vignettes) was sufficient to make their judgments more severe when compared to the judgments of the unhypnotized. For Wheatley and Haidt, the above experiment suggests that “an arbitrarily induced gut-level response (disgust) would be *used as information* for moral judgment [emphasis added]” (Wheatley and Haidt 2005, 780). And as Prinz argues, this experiment suggests that “emotions can influence moral evaluations even when the emotions are induced by morally irrelevant factors” (Prinz 2007, 28). In sum, if emotion constitutes (at least in part) moral judgments, it explains why the emotional experience itself is sufficient to trigger a moral judgment. However, as we can already see that though the influence claim is sympathetic towards Prinz’s claim that moral judgment comprises emotion, the latter does not necessarily follow from the former. As Prinz argues, “... empathy for the victim of a crime could instill the desire for punishment, and that could lead us to weigh evidence selectively in assigning blame to a suspect. On this view, the assignment of blame would not need to be an emotional judgment in its own right, even though emotions played a role in bringing it about. Emotions play a causal role, here, but they are not constitutive” (Prinz 2007, 28). The point here is that non-EAMs can devise various explanations to explain the influence claim. This claim by itself would appear to be insufficient on its own to support EAMs. But when we couple the influence claim with the co-occurrence claim, such empirical claims help us build a case for EAMs. EAMs can easily explain such phenomena. Let us continue to build the EAM case with the next empirical claim.

Emotions are sufficient for moral judgments. Prinz argues that “having a moral attitude is a matter of having an emotional disposition. If this is right, then someone should be able to have a moral attitude in the absence of any rational justification. Emotional attitudes should be sufficient for moral attitudes” (Prinz 2007, 29). Prinz’s argument here is that moral judgment (as a psychological state) is a matter of having an emotional disposition. The sufficiency claim supports Prinz’s EAM

in the sense that the latter can explain the former. He also argues that there is reason to believe that the sufficiency claim is true. To support this he proffers the observation of moral dumbfounding by Haidt et al where “[moral] judgment preceded reasoning, judgments were based on gut feelings than on reasoning, and participants more frequently laughed and directly stated that they had no reasons to support their judgments” (Haidt, Bjorklund and Murphy 2000, 1). In this experiment, a group of thirty participants were presented with multiple scenarios, and one of them is as follows:

“Julie and Mark, who are brother and sister are traveling together in France. They are both on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At very least it would be a new experience for each of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy it, but they decide not to do it again.’ They keep that night as a special secret between them, which makes them feel even closer to each other. So what do you think about this? Was it wrong for them to have sex?” (Haidt, Bjorklund and Murphy 2000, 15).

The participants were then asked to judge whether these actions were morally wrong. The experimenters would subsequently ‘argue’ with them, so to undermine the reasons given for their judgments. In terms of the kinds of arguments that the participants would give for their judgments, the following are Haidt et al’s observations: (a) “participants [often] made ‘unsupported declarations’, e.g., ‘It’s just wrong to do that!’ or ‘That’s terrible!’”, (b) participants often “thought [the] action was wrong but they could not find the words to explain themselves”, (c) “participants often said ‘I don’t know,’ sometimes several times in a row”, and (d) “participants would start giving an argument but as they were talking they realized that the argument was not going to work and they stopped in the middle of it, without any prompting from the experimenter” (Haidt, Bjorklund and Murphy 2000, 9). One plausible explanation for these observations is that the emotion itself is sufficient for moral judgment. As odd as it sounds, it appears in this experiment that the judgment ‘incest is wrong’ is based on the reason that it is ‘just wrong’. EAMs can explain this ‘just wrong’ phenomena. If the term ‘wrong’ in the utterance ‘incest is wrong’ expresses a noncognitive state of disapproval (e.g.), this claim can explain why we appear to make such judgments and maintain our stance regardless of what arguments against this stance are thrown at us. To say that it is ‘just wrong’ can mean that I just disapprove of it, period.

Emotions are developmentally necessary for moral judgments. Mark Johnston argues that: "Seeing the utterly specific ways in which a situation... is appealing or repellent requires an appropriate affective engagement with the situation... Absence of appropriate affect makes us aspect-blind... If one has never been moved or affected by the determinate ways in which things are beautiful or charming or erotic or banal or sublime or horrific or appealing, then one is ignorant of the relevant determinate values" (Johnston 2001, 181-182). To put Johnston's argument in our context, we can understand his argument to imply that without first having the emotional capacity to be horrified and experiencing the horrific (e.g.), we cannot genuinely make judgments like *x* is horrific. Johnston's argument does not appear to confine itself to moral judgments. The beautiful, charming, and erotic (e.g.) are not necessarily moral values. However, Johnston's argument remains applicable. The reason for this is that while he does not confine his argument to moral values, his argument at the very least encompasses such values. Now if we put aside Johnston's metaphysical concerns, we can understand the developmental claim as follows: without certain emotions, we are unable to make genuine moral judgments. This position is similar to that of Michael Slote. He argues that "people incapable of empathy are... incapable of genuinely approving or disapproving the virtues and vices of others" (Slote 2010, 36).

Now if this developmental claim has empirical support and EAMs can explain the phenomena revealed by this claim, the developmental claim would in this sense support EAMs. Let us start with the empirical support. Prinz argues that psychopaths who suffer from emotional deficits "seem to comprehend morality, but they really don't. They use moral terms in a way that deviates strikingly from the way non-psychopaths use those terms. These deviations suggest that they do not possess moral concepts; or at least their moral concepts are fundamentally different from ours" (Prinz 2007, 43). Prinz relies on a series of empirical studies to support this argument. However, for our purposes, it would suffice to consider one example. He cites Hervey Cleckley who argues that:

"The [psychopath] is unfamiliar with... what might be called personal values... Beauty and ugliness, except in a very superficial sense, goodness, evil, love, horror, and humor have no actual meaning, no power to move him.... It is as though he were colorblind, despite his sharp intelligence, to this aspect of human existence. It cannot be explained to him

because there is nothing in his orbit of awareness that can bridge the gap with comparison. He can repeat the words and say glibly that he understands, and there is no way for him to realize that he does not understand” (Cleckley 1941, 40).

In other words, a psychopath is like a person who is colour blind. Just as a colour-blind person cannot grasp concepts like red, green and yellow like non-colour-blind persons, a psychopath cannot grasp concepts like good and evil like non-psychopathic people who do not have emotional deficits. Considering this, if moral judgment (as an utterance/proposition) expresses emotion or have emotion as its referent (e.g.), these can explain why people with emotional deficits cannot make genuine moral judgments. It appears that to genuinely use moral terms like ‘right’ and ‘wrong’ in moral judgments, one must have (or at least have had) certain emotional capacities.

In sum, the case for EAMs become stronger with the introduction of each empirical claim. The co-occurrence claim can support EAMs. But there is quite a lot of wiggle room to accept this claim while providing a non-EAM explanation for it. Emotion may co-occur with moral judgments. But this claim does not even imply the claim that ‘emotion causes moral judgment’ let alone the claim that ‘moral judgment (as a psychological state) is constituted by emotion’ (e.g.). The influence claim can also support EAMs. This claim goes beyond claiming that there is a correlation between emotion and moral judgment and establishes a kind of causal tie between them. Emotion (when present) can cause a judgment to be more severe when compared to a situation where the emotion is absent. As we have seen, EAMs can explain the influence claim. The influence claim however falls short of claiming that emotion can cause a moral judgment *simpliciter*. This is where the sufficiency claim comes in, i.e. emotion can cause a moral judgment *simpliciter*. EAMs can explain this. Lastly, we see with the developmental claim that without certain emotional capacities we cannot make genuine moral judgments. Unlike the previous empirical claims, the developmental claim establishes a stronger nexus between emotion and moral judgment in the ‘no emotion = no moral judgment’ sense. EAMs’ explanatory power also extends to this claim. Now if the number of such empirical claims pile up and EAMs can explain them, EAMs’ explanatory power over these phenomena is a reason for accepting that EAMs are a set of plausible accounts of moral judgments. With this in mind, let us proceed to the next section of this chapter outlining the various token EAM theories.

Emotivism

Michael Smith argues that for emotivists “moral judgments express and arouse emotions, not beliefs. Saying that an act is right or wrong was thus supposed to be rather like saying ‘Boo!’ or ‘Hooray!’” (M. Smith 1998). And with more detail Mark van Roojen argues that for emotivists:

“... moral terms in grammatically assertive utterances function to express emotion and perhaps also to elicit similar emotions in others... [T]he right way to explain the meaning of such terms is to point out that they are conventional devices for performing a certain sort of speech act, one which if sincere requires that the speaker have a certain attitude. Sentences employing general predicates of negative evaluation such as ‘wrong’, ‘bad’, and ‘vicious’ signal negative non-cognitive attitudes. Thus to call a person virtuous is to express an attitude of approval and the speech act of doing so is analogous to the speech act performed when we cheer for that person” (van Roojen 2018).

The point that I want to draw out is that it appears that the utterance that x is wrong expresses the psychological state of judging and this state is an emotion. Specifically, it is the term ‘wrong’ in ‘ x is wrong’ that express the negative emotion. This reveals how emotivism is a token EAM theory. Emotivists would appear to accept the claim that certain emotions are necessarily involved in the meaning of the utterance that x has a moral attribute. With this in mind, I now consider the position of one prominent emotivist, i.e. Charles Stevenson.

Stevenson argues that “When you tell a man that he oughtn’t to steal, your object isn’t merely to let him know that people disapprove of stealing. You are attempting, rather, to get him to disapprove of it. Your ethical judgment has a quasi-imperative force which operating through suggestion, and intensified by your tone of voice, readily permits you to begin to influence, to modify, his interests” (Stevenson 1937, 18-19). He also argues that “a moralist is so often a reformer is scarcely an accident. His judgments plead and advise, and open the way to counteradvice. In this way moral judgments go beyond cognition, speaking to the conative-affective natures of men” (Stevenson 1944, 13). Now let us investigate Stevenson’s argument further. First, with the above we can say that moral judgment refers to the uttered proposition (what I *tell* another person) that ‘stealing is wrong’. What Stevenson means by moral judgment here is not the psychological state of judging that x has a moral attribute. Rather, it appears that he is referring to the utterance that

x has a moral attribute. Second, if we grant this understanding of moral judgment, we can understand Stevenson's argument this way: moral judgments (a) express what one approves/disapproves of, and (b) recommend what one approves/disapproves of. To explain this ability of moral judgments to 'plead and advise', he argues that the "emotive meaning of a word is the power that the word acquires, on account of its history in emotional situations, to evoke or directly express attitudes, as distinct from describing or designating them... In virtue of this kind of meaning, ethical judgments alter attitudes, not by an appeal to self-conscious efforts... but by the more flexible mechanism of suggestion" (Stevenson 1944, 33). In other words, moral judgments recommend as they contain emotive terms that can evoke and alter attitudes. Now let us probe deeper into this apparent connection between emotions and emotive terms used in moral judgments. For Stevenson, emotive terms are like interjections. And he argues that "interjections... are *like* sighs, shrieks, groans... they can be used to 'give vent' to the emotions... in much the same way. The word 'hurrah,' for instance, serves much the same purpose as any simple cry of enthusiasm, and releases the emotions with equal directness" (Stevenson 1944, 38). With the above, it appears the emotive term 'bad' is like the interjection 'boo'. Just as 'boo' expresses my displeasure at what someone did or say (e.g.), the use of the term 'bad' in moral judgment can be an expression of my indignation (e.g.) toward an action. With this, I think we can see how Stevenson fits the bill as a token EAM theory. For Stevenson, certain emotions are necessarily involved in the meaning of the term 'bad' (e.g.) in the moral judgment that ' x is bad'.

Stevenson's theory (as a representative of emotivism) does well in explaining the apparent connection between moral judgment and motivation. If we sincerely judge that 'murder is bad', we are usually disinclined to murder someone. We can explain this disinclination with the claim that the use of the term 'bad' in the judgment that 'murder is bad' expresses my abhorrence (e.g.) toward murder. This is because if our moral judgment expresses our abhorrence toward murder, we would likely be disinclined toward it. This being said, Stevenson's theory is wanting in many respects. I will highlight two criticisms. First, Stevenson's theory does not appear to adequately explain rational arguments in moral discourse. Stevenson appears to ignore what Simon Blackburn calls "the propositional surface of ethical discourse" (Blackburn 2002, 127). For Stevenson, when I judge that ' x is bad', it is an expression of my emotion toward x and *not* an expression of my belief that ' x is bad'. My moral judgment that ' x is bad' would as such not be the sort of thing that can be true or false. Stevenson's theory does not appear to adequately explain why we often argue about

moral judgments in terms of their truth and falsehood. As Smith argues, “since [emotivism] entailed that moral judgments elude assessment in terms of truth and falsehood, it suggested that rational argument about morals might be at best inappropriate, and at worst impossible” (M. Smith 1998). Second, Stevenson appears to hold the position that the emotion (linked to the emotive term used in the moral judgment) must be ‘online’ whenever one makes a moral judgment. As Shaun Nichols argues, “emotivists maintained that a person must actually have the emotion that he is expressing when he utters a moral condemnation” (Nichols 2008, 256). This position appears to be untenable. It appears that we can judge that ‘Ted Bundy is bad’ without feeling indignant toward him or his actions.

I highlight the above two criticisms to segue into how EAMs developed since then to explain certain features of morality. More precisely, I highlight these criticisms as a tool to segue into two emotivist-inspired theorists, i.e. Blackburn and Allan Gibbard. Both are known as expressivists and we can understand their theories as a development on emotivism. I intend to show in the next section how Blackburn builds on emotivism and in particular how he addresses the first criticism. I would do the same with Gibbard but with a mind of showing how he addresses the second criticism.¹⁷ Considering this, let us now proceed to the next section.

Expressivism

Expressivism builds on what emotivism does well, including the latter’s ability to explain the connection between moral judgment and motivation. As van Roojen argues on behalf of expressivists:

“[T]he function of moral language is to express desire like attitudes. The fact that moral language does so is supposed to explain the intuitively tight connection between moral opinion and action – that people’s actions provide good evidence about the morality they accept... Expressivists and emotivists agree that simple indicative moral sentences are conventional devices for the expression of pro and con attitudes as opposed to cognitive

¹⁷ To avoid doubt, I am not claiming that Blackburn’s theory addresses only the first criticism and not the second. Neither am I claiming that Gibbard’s theory addresses only the second criticism and not the first. What I intend to do is to consider (a) how Blackburn’s theory addresses the first criticism as a token EAM theory, and (b) how Gibbard’s theory addresses the second criticism as a token EAM theory.

attitudes such as belief. Contemporary expressivists have not repudiated emotivism; rather, they have developed it" (van Roojen 2018).

In applying the above to my thesis, I make three points. First, expressivists and emotivists are EAMs in the sense that certain emotions are necessarily involved in the meaning of the moral judgment that 'x is bad'. Second, both expressivists and emotivists are also committed to noncognitivism in that they reject the claim that a moral judgment that 'x is bad' (e.g.) expresses a belief that 'x is bad'. Rather, moral judgments are expressions of some noncognitive state. Third, expressivists continued to build on the 'structure' initially built by the emotivists. The expressivist does not strip the existing emotivist structure. Rather, the expressivist builds on the structure with a sophistication that the emotivist lacked. The question that can be asked now is *how does expressivism 'build on' emotivism*. I would answer this question considering Blackburn and Gibbard. Let us start with Blackburn and his quasi-realist position.

Blackburn argues that "[when we say that something is good or right...] [w]e avow a practical state. 'Avowal' here means that we express this state, make it public, or communicate it. We intend coordination with similar avowals or potential avowals from others, and this is the point of the communication. When this coordination is achieved, an intended direction is given to our joint practical lives and choices" (Blackburn 1998, 68-69). Here we see that the judgment (as an utterance) that x has a moral attribute expresses a practical state.¹⁸ We can understand what Blackburn means by this in the following way: to genuinely say *I believe that X is good or right* is to express "an approximately favourable valuation of X", to genuinely and explicitly question whether X is good or right is to express one's "wondering [of] what to do/what to admire or value", to genuinely and explicitly deny that X is good or right is to express one's rejection of "a favourable attitude to X", to genuinely say *I am certain that X is good or right* is to have "a settled attitude/[rejection of] the possibility that improvement could result in change" (Blackburn 1998, 70).¹⁹ These practical states

¹⁸ I note that Blackburn scarcely uses the term 'moral judgment' in *Ruling Passions*. Given that emotivists use the term 'moral judgment' to refer to a moral utterance or a moral term within a moral sentence, I shall use the term 'moral judgment' in the context of Blackburn's expressivism to refer to the same sort of things. I do this to avoid terminological confusion, especially as I want to present Blackburn's quasi-realism as an EAM.

¹⁹ Blackburn argues that "[s]aying that something is good when we do not really value it is either deceiving others about our state, or is the result of self-deception" (Blackburn 1998, 69). This coupled with his argument above makes it appear that for Blackburn it is only when we genuinely say that x is good that we express a practical state. Now what does it mean to *genuinely* say that x is good? It seems that this genuineness is a correspondence between *saying* that x is good and *thinking* that x is good. If I do not think that x is good but yet say that x is good, then the

involve what Antti Kaupinnen argues to be “higher-order attitudes towards desires and preferences” (Kaupinnen 2016). This means that to judge that ‘*x* is bad’ is not only to express a con-attitude toward *x*, but also a pro-attitude toward those who have a con-attitude toward *x* and a con-attitude toward those who have a pro-attitude toward *x*. Blackburn’s argument is that we start off with “simple preferences” toward *x*. Taking it to the next level, there can be “hostility” toward *x*. The next level up would be hostility to those who do not share our hostility toward *x*. (Blackburn 1998, 9).

Now comes the question of how Blackburn reconciles moral judgments as expressions of appetitive states with the obvious representational appearances of such judgments. Blackburn argues that “[o]n the one hand [ethical commitments] seem to be representative. We talk of ethical knowledge, and truth and reason. On the other hand they seem to be appetitive. Ethical knowledge motivates us, some of the time, and feelings like anger, guilt, shame and indignation give us a central component of our practical stances” (Blackburn 2002, 124). It appears that the difficulty for Blackburn here lies in reconciling the underlying practical states of moral judgments (which are not the sort of thing that can be true or false) with moral propositions or discourse. A proposition appears to be the sort of thing that can be true or false. And discourse often involve an underlying assumption that moral propositions discussed can be true or false). As we can see, Blackburn takes the extra step (that emotivists do not) to reconcile the appetitive feature that underlies moral judgments with their propositional appearance. Blackburn appears to argue that moral judgments are expression of practical states while maintaining that (a) such judgments with their underlying noncognitive practical states remain compatible with moral discourse, and (b) discourse about such judgments remains appropriate.

On how moral judgments (as described above) remain compatible with moral discourse, part of the answer lies in Blackburn’s account of how we can make sense of claims like “‘*x* has a moral attribute’ is true’. His argument is:

latter saying is disingenuous or self-deceptive. If this is correct, it appears that we can also say that the *thinking of x as good* is the practical state of having a favourable attitude toward *x*.

“To worry whether a moral judgement is true is to worry which attitude to take towards something, and ... to say that a moral judgement is true is to repeat the judgement. To claim knowledge of a moral judgement is to claim a standpoint such that no improvement will lead to a reversal of attitude. I know that child abuse is a bad thing, and in saying that I express my conviction that no further facts wait to come in. There is nothing further about human life or children’s lives that might come along and upset that verdict” (Blackburn 2002, 128-129).

What I take from this and Blackburn’s overall moral philosophy is that when we expressly or implicitly say that a moral judgment is true, we are not saying that this moral judgment corresponds with any fact out there. Rather to say or imply that “‘x has a moral attribute’ is true” is just another way of repeating the judgment ‘x has a moral attribute’. If we take as a given that the claim “‘x has a moral attribute’ is true” does not correspond with any facts out there, there appears to be no need to accept that moral judgments are expressions of cognitive states like beliefs. And if we take as a given that the claim “‘x has a moral attribute’ is true” is just a repetition of the judgment ‘x has a moral attribute’ and this judgment expresses one’s attitude toward x, there also appears to be no incompatibility between the claims that (a) moral judgments are expressions of noncognitive practical states and (b) moral judgments can be true or false. In sum, Blackburn’s position is that though moral judgments are expressions of noncognitive practical states there is no need to refrain from talk about the truth of such judgments.

On how moral discourse remains appropriate despite moral judgments being expressions of noncognitive practical states, Blackburn starts by rejecting the emotivist position of some that “regretted the propositional surface of moral discourse” (Blackburn 2002, 127). He rejects the position that it would be better to display moral judgments solely in terms of pro- and con-attitudes. Blackburn argues against this position by saying that “the propositional surface [of moral judgments] is both explicable, and justifiable”. Their propositional surface is explicable as it provides us with an array of resources to “discuss, accept, reject, attitudes...” so to determine which “networks of attitudes and beliefs” stands up to scrutiny. Their propositional surface is also justifiable as expressing our “simple ‘boo-hooray’ language” in propositions is the only expression through which we can “communicate and argue as necessary” (Blackburn 2002, 127-128). To sum up the above in one sentence, for Blackburn moral discourse remains appropriate (despite their

noncognitive underpinnings) on grounds that we need it. The question that we can ask now is what do we need moral discourse for? Blackburn's answer can be roughly and briefly put in the following manner: if we rely on 'simple boo-hooray' language alone, it would inevitably lead to violent conflict. As Blackburn argues: "If we tried an ethical life modelled on [simple boo-hooray language], it would risk unpredictability and injustice... Coordinating our attitudes [via language] is one of the most important tasks that face us when, as with ethical conflict, the result of failure may be nothing short of conflict and death" (Blackburn 2002, 128).

With the above, I hope the reader can see how Blackburn's quasi-realism builds on emotivism by addressing the criticism that emotivism cannot make sense of moral discourse. And given that I intend this criticism as a segue into a discussion of Blackburn's quasi-realism and how it can fit as a token EAM theory, allow me to sum up that for him (as seen above) certain practical states (emotions included) are necessarily involved in the meaning of moral judgments (as genuine utterances or propositions) that x has a positive/negative moral attribute.

Now I will turn my attention to how Gibbard builds on emotivism by addressing the criticism that emotivism requires emotions to be online when making moral judgments. Before elaborating further on this criticism, I would like to make clear what we mean here by moral judgment. Gibbard argues that "moral judgments concern which moral sentiments are warranted or justified... To call a feeling warranted... is (roughly) to express one's acceptance of norms that permit the feeling. Moral judgments, then, *express a state of mind* that is not, in the strictest sense, belief in a moral fact [emphasis added]" (Darwall, Gibbard and Railton 1992, 150). As it appears, the text hints toward moral judgments as *expressions* of a psychological state (not the psychological state itself). From this, we could understand Gibbardian moral judgments as utterances (or verbalised propositions) that express a psychological state. This, however, does not preclude us from holding that moral judgments for Gibbard can also be a (kind of) psychological state. Stephen Darwall, Gibbard and Peter Railton argue that "[e]motivists hold that a moral judgment consists in a feeling – or better, in a disposition to have certain feelings. It seems, though, that a person can judge something wrong even if he has lost all disposition to feelings about it" (Darwall, Gibbard and Railton 1992, 149). With the use of the word 'consist' to signify the relationship between moral judgment and emotion, this seems to suggest that moral judgment is/involves a psychological state.

This interpretation gains further support when earlier in the same text, Darwall, Gibbard and Railton argue that expressivists “explain moral judgments as something other than beliefs” (Darwall, Gibbard and Railton 1992, 148). This comparison of moral judgments with belief (a psychological state) seems to suggest that moral judgments are/involve a psychological state. I would, however, not go down this path so to be consistent with the earlier discussion on moral judgments as utterances/propositions.²⁰

Now with the above in mind, let us consider in more detail the criticism that Gibbard’s position addresses. The passage quoted at the start of the preceding paragraph stresses an apparent expectation that should be met by any plausible EAM, i.e. EAMs must be able to accommodate circumstances where one can genuinely claim that x has a moral attribute without feeling any emotion at all. And as we have seen earlier, Nichols argues that emotivists do not meet this expectation. I also hinted earlier that Gibbard appears to meet this expectation. How then does Gibbard do so? He argues that “[t]o feel guilt or resentment is not in itself to make a moral judgment. A person can feel guilty and yet think he has done no wrong. He then thinks it makes no sense to feel the way he does, that his feelings are irrational ... moral judgments are not feelings but judgments of what moral feelings it is rational to have. Feelings... can be apt or not, and moral judgments are judgments of when guilt and resentment are apt” (Gibbard 1990, 6). In other words, we can say that, for Gibbard, to claim that x has a negative moral attribute is also to claim that it is apt to feel guilt and/or resentment toward x . Now if so, what does this judgment ‘it is apt to feel guilt and/or resentment toward x ’ express? The short answer is that this judgment expresses a noncognitive psychological state. Here is the longer answer. Gibbard argues that “... to call something rational [or apt] is to express one’s acceptance of a system of norms that permits it – the act or belief or feeling” (Gibbard 1993, 34). Now what is this state of norm acceptance? He argues that it is a “basic kind of noncognitive state, an evolutionary adaptation for linguistically achieved coordination that is not analyzable in terms of other attitudes. It is non-cognitive, because it is essentially a motivational tendency to act or feel in ways that we are prepared to avow in discussion about what to do” (Kauppinen 2016).²¹ In Gibbard’s own words:

²⁰ I would discuss moral judgments as psychological states more extensively in the sections titled *Dispositionalism* and *Pluralism*.

²¹ In *Thinking How to Live*, Gibbard revised to an extent his position on norm acceptance. However, given that the revision involved does not put his theory out of the expressivist framework and is not crucial to the aim of this chapter to present an overview of EAMs, I chose not to discuss this revision.

The “state of accepting a norm... is identified by its place in a syndrome of tendencies toward action and avowal- a syndrome produced by the language-infused system of coordination peculiar to human beings. The system works through discussion of absent situations, and it allows for the delicate adjustments of coordination that human social life requires. The syndrome that manifests accepting a norm takes in normative discussion and normative governance. In this normative discussion, in unrestrained contexts, one tends to avow the norm. One tends to be influenced by the avowals of others, and to be responsive to their demands for consistency. Normative governance by the norm is a tendency to conform to it. *Accepting* a norm is whatever psychic state, if any, gives rise to this syndrome of avowal of the norm and governance by it” (Gibbard 1990, 75).

With the above passage, the point I would like to emphasize here is that like emotivists, moral judgments for Gibbard express a noncognitive state. The point of departure for Gibbard from the emotivists is that instead of arguing that moral judgments express noncognitive emotions, he argues that moral judgments express the noncognitive psychological state of accepting the norms that govern the elicitation of such emotions.

At this point, I believe the reader can already see how Gibbard’s account of moral judgment accommodates circumstances where one can genuinely claim that x has a moral attribute without feeling any emotion at all. There is no need for any emotion to be online when one makes a moral judgment. I do not, for example, have to feel resentment toward x when I genuinely claim that x has a negative moral attribute. When I claim that x has a negative moral attribute, I only need to feel that it is apt for me to feel resentment toward x . As Nichols argues, “... even if one has lost any disposition to feel guilty about a certain action, one can still think that feeling guilt is *warranted*. Thus the problem of absent feeling is addressed” (Nichols 2008, 258). And given that EAMs should be able to accommodate such instances of ‘cool’ moral judgments but emotivism cannot, we can consider Gibbard’s expressivism as an improvement on emotivism in this respect. As I intended this emotivist problem as segue into a discussion of Gibbard’s expressivism and how it can fit as a token EAM theory, I sum up (before moving to the next section) that for Gibbard the noncognitive state of accepting the norms that govern the elicitation of resentment and guilt toward x is necessarily involved in the meaning of moral judgments (as genuine claims or propositions) that x has a negative moral attribute.

Empathy-based Sentimentalism

We saw with the above that for some EAMs moral judgments express emotions/non-cognitive practical states (involving emotions)/noncognitive states of accepting norms governing emotional responses. Now we turn to a token EAM theory where we can explain the nature of moral judgment in terms of its connection to our empathic capacities. The theory that I have in mind is that of Slote. As we shall see, Slote's position is that certain empathic attitudes of moral approval and disapproval are necessarily involved in moral judgments.

Slote starts by arguing that moral approval (as a psychological state) is constituted by a second order empathy: "if agents' actions reflect empathic concern for (the well being or wishes of) others, empathic beings will feel warmly or tenderly toward them, and such warmth and tenderness empathically reflect the empathic warmth or tenderness of the agents... such empathy with empathy... constitutes moral approval... for agents and/or their actions" (Slote 2010, 34-35). As we can see, this empathic warmth is not a first-order empathy. It does not occur when an empathic agent empathises with the needs of others. Instead, empathic warmth is a second-order empathy. It occurs when an empathic agent empathises with the feelings and motivations of another empathic agent who empathises with the needs of others. Inversely, as Slote argues, moral disapproval can "be understood on analogy with [moral] approval. If a person's actions toward others exhibit a basic lack of empathy, then empathic people will tend to be chilled (or at least 'left cold') by those actions, and... those (reflective) feelings toward the agent constitute moral disapproval" (Slote 2010, 35). In other words, moral disapproval is empathic chill. And empathic chill occurs when an empathic agent feels the cold lack of concern that (apparently) non-empathic agents have toward the needs of others.

Now that we considered moral approval/disapproval, let us now consider what for Slote are moral judgments. In Slote's chapter titled *Empathy in Moral Judgment*, he appears to use the terms moral judgments, claims, sentences, utterances interchangeably (Slote 2010, 45-56). We can take this interchangeability to mean that there is some common element that unites claims, sentences and utterances. This common element in turn can provide us with a clue to what Slote means by 'moral

judgment'. We can understand claims, sentences and utterances as signifiers of a psychological state. The claim/sentence/utterance that 'x has a moral attribute' can signify (in the sense that it can 'point toward') the psychological state that corresponds to this claim/sentence/utterance. In short, moral judgments for Slote can be psychological states that is connected in one way or another to states of moral approval and disapproval. This being said, we can also understand moral judgments as *propositions* that x has a moral attribute (as opposed to the state of judging that x has a moral attribute). What I mean here by proposition is simply a statement that can be true or false. This statement can be communicated via the linguistic vehicle of a sentence and expressed through an utterance. This view of moral judgments as propositions can help explain Slote's interchangeability of claims, sentences and utterances. If we understand moral claims as propositions, we can see moral sentences and utterances as forms of expressions of moral claims. Slote's reference to such judgments as "explicit moral judgments" (Slote 2010, 43) can also support the view of moral judgments as propositions. It seems plausible for explicit moral judgments in this context to mean 'moral judgments that are clearly communicated in words'. This is especially so when understood as a precursor to his argument for a "theory of the meaning of moral terms" (Slote 2010, 26). Though there appears to be at least two types of Slotean moral judgments, I narrow my discussion in this section to moral judgments as propositions. I do this not because moral judgments as psychological states are unimportant. Rather, I do so to continue on from my earlier discussion in this chapter on moral judgments as propositions/utterances.

With the above in mind, we are now able to ask *what is the connection between moral approval/disapproval and moral judgments (as propositions)*. We can describe Slote's short answer as follows: moral approval/disapproval fixes the reference of moral terms within a moral judgment in such a way that moral terms like good or bad refer to whatever causes moral approval/disapproval. We can see Slote's longer answer in what he calls "semi-Kripkean reference fixing approach" (Slote 2010, 49). He first describes Kripke's approach as follows: "... if Kripke explicates (objective) 'red' by saying that its reference is fixed a priori by our subjective experience of redness... then we can proceed analogously with 'right' or '(morally) good'. We can say... that the reference of 'right' is fixed a priori by the subjective feeling of warmth we feel in regard to certain actions..." (Slote 2010, 59). Slote then builds on this Kripkean approach by arguing that "... what fixes... the reference is not some thin, subjective notion of experience but, rather, certain experiences *as directed toward others and as reflecting what others feel...*" (Slote 2010, 64). Now Slote's

reference-fixing theory differs (e.g.) from emotivism and expressivism in at least two respects. First, for emotivists/expressivists, moral terms in moral judgments express various noncognitive psychological states themselves. But for Slote, he argues that “[m]oral claims don’t describe our (tendencies toward) attitudes of approval or our (tendencies toward) empathic reactions of warmth or chill; rather, these attitudes and empathic reactions (are used to) fix the (objective) reference of moral terms and utterances” (Slote 2010, 67). Hence, Slote’s account is not one where moral judgments express emotion. Rather, his account is one where the referents of moral terms within moral judgments are fixed by emotion. This reference-fixing allows Slote to shift the focus of moral judgments from an internal psychological state to the intentional and external object of the internal psychological state. As Slote argues, “the present reference fixing form of sentimentalism regards moral claims as *not* being about our observer reactions to acts and agents, and sees them, rather as being primarily *about the acts or agents we react to...*” (Slote 2010, 67). Moral terms within moral judgments, hence, refer not to moral approval/disapproval, but to whatever external object that causes the internal psychological state of moral approval/disapproval. It is in this sense that moral terms are fixed by moral approval/disapproval. Here we see that there can be a distinction between what moral terms *refer to* and what these referents *are about*.

Second, for emotivist/expressivists, moral judgments are not the sort of thing that can be true or false (at least in a correspondence to fact manner). But for Slote, moral judgments are the sort of thing that can be true or false (in a correspondence to fact manner). Slote argues that “[i]f empathy derived agent directed warm feelings are crucial to the reference fixing of ‘right’ or ‘good’, then, given our judgment free sentimentalist conception of moral approval, we can also say that moral approval plays a major role in fixing the reference of that term. What is right (or morally good) is what causes or tends to cause nonjudgmental approval... moral rightness, wrongness, and goodness are real properties of agents and their acts, attitudes, and motives” (Slote 2010, 66-67). Reading this point alongside the preceding paragraph, Slote is arguing that the meaning of the proposition ‘*x* is good’ is that *x* has the property of goodness which causes or tends to cause moral approval. This proposition that ‘*x* is good’ can as such turn out to be true or false (in a correspondence to fact manner).²²

²² I note that I highlight the differences between emotivism/expressivism and Slote’s theory not with the intention of pitting one theory against the other. Nor do I highlight them as a prelude to deciding between them. Rather, I highlight these differences to emphasize the diversity of EAMs. Though some may find a token theory unpalatable

In sum, the moral judgment/*proposition* that 'Mother Teresa is good' can be expressed verbally in the form of a *sentence* and orally in the form of an *utterance*. The term 'good' in this moral judgment/*proposition* *refers to* whatever causes moral approval. This moral approval is the agent's experience of empathic warmth, where the agent feels empathically warm toward Mother Teresa who is empathically warm toward the needs of others. And the agent's empathic warmth *is about* Mother Teresa's actions and underlying empathic warmth toward Kolkata's poor. To put these in one sentence, certain terms in moral judgments as propositions (expressible via sentences and/or utterances) have empathic warmth/chill (and what they are about) as referents.

Dispositionalism

So far, we discussed emotivism, expressivism and empathy-based sentimentalism. We saw how these token EAM theories vary in their explications on the connection between emotion and moral judgment. For emotivists/expressivists, moral judgments express emotions. For empathy-based sentimentalists like Slote, moral judgements have emotions as referents. For the former, moral judgments are not truth apt (at least in the correspondence to fact manner, if at all). For the latter, moral judgments are truth apt (in the correspondence to fact manner). This difference between them can serve as a segue into dispositionalism. Joshua Gert argues that "[d]ispositionalism... makes moral judgements obviously and unproblematically truth-apt by giving them straightforwardly descriptive truth conditions. Such views still count as sentimentalist because those truth conditions are facts about human sentiments. According to dispositionalism, being morally wrong is a matter of actually eliciting the relevant attitude in the relevant class of people" (Gert 2015). As we can see, for the dispositionalist, whether '*x* is wrong' is true depends on whether a 'relevant class' elicits a 'relevant attitude' toward *x*. With this in mind, the prominent dispositionalist that I have in mind is Prinz and I shall in this section outline his theory.

Prinz argues that: "[w]hen we say, 'That's wrong!' we convey our feelings and also aim to assert a fact. 'Pickpocketing is wrong' represents the fact that pickpocketing has the property of wrongness.

for whatever reason, there is a likelihood that they would find another more palatable token theory in light of EAMs' diversity.

Moral wrongness... is the property of being the object of disapprobation" (Prinz 2007, 100). This claim requires some unpacking. The moral judgment involved here is the utterance that 'pickpocketing is wrong'.²³ If so, for Prinz moral judgments express our feelings. In saying that 'x is wrong' we convey our negative feelings. Such judgments do not express emotions strictly so called (e.g. anger, disgust, guilt, etc.). Rather, they express a *sentiment* which for Prinz is a psychological "disposition whose occurrent manifestations (or working memory encodings, or neural activation patterns) are emotions" (Prinz 2007, 84). This means that when I genuinely utter that 'x is wrong' I am expressing my disposition "to feel other-blame emotions when we are victims of transgressions [of the kind of x] and self-blame emotions when we transgress [by committing acts of the kind of x]" (Prinz 2007, 96). Prinz also provides us with a detailed example of how an external event activates a sentiment which in turn activates an emotion:

"First, the observer interprets pickpocketing as an instance of stealing. This is a categorization stage; it involves drawing on prior knowledge to classify an observed action. Stealing is something toward which the observer has a moral sentiment in long-term memory. His sentiment toward stealing constitutes a rule. When the idea of stealing enters his mind, the rule causes the sentiment to become active. This is a rule retrieval stage. At this stage, the sentiment is not yet experienced as an emotion. Sentiments are dispositional. Once a sentiment is activated, contextual factors are used to determine which emotion will be elicited. Because the observer is not the author of the action, and because stealing is an autonomy norm, the elicited emotion is anger. This is an emotion elicitation stage... The anger was triggered by the experience of pickpocketing, and it gets bound to the representation of pickpocketing. The result is a compound state: anger at pickpocketing" (Prinz 2007, 96).

Now as we can see via the stages of categorisation, rule retrieval and elicitation, Prinz's dispositionalism is united with emotivists and expressivists in its commitment to the claim that moral judgments express an underlying psychological state. Prinz's point of departure is his

²³ Strictly, the utterance that 'pickpocketing is wrong' is not for Prinz a moral judgment. Rather, it is the verbalized form of an underlying psychological state. As Prinz argues, "[w]e often talk as if verbalizations of moral judgments were moral judgments in their own right. I will refer to sentences, such as 'Pickpocketing is wrong' as a verbalized moral judgment. But this label is really shorthand. 'Pickpocketing is wrong' is not a judgment; it's a string of words. In calling it a verbalized moral judgment, I mean it is a verbal form that might be used to express a moral judgment" (Prinz 2007, 100). This being said, in this section, I would refer to Prinz's verbalized moral judgment as moral judgments in the 'utterance' sense to ensure that there is no major shift in the meaning of moral judgment from the earlier sections. I would then outline his position on moral judgment as a psychological state.

special use of the term 'sentiment' as a psychological *disposition* to feel certain emotions. If moral judgments express such a disposition, there is no need for emotions to be online when we make such judgments. For Prinz, "dispassionate moral judgments" are possible (Prinz 2006, 37). He argues that "[i]f you cross me, I may experience anger, and that anger qualifies as a token of my concept WRONG. It is the experience of anger that alerts me to the fact that you have done something morally questionable, and I give voice to the anger when I judge that you were wrong to treat me that way. [Such] ordinary moral judgments are hot" (Prinz 2006, 37-38). But this need not be the case whenever I claim that 'x is wrong'. There are 'cool' moral judgments and as Prinz argues, "[s]uch judgments are not immediately motivating, but they express a dispositional state that would motivate under the right circumstances" (Prinz 2006, 38). In sum, the judgment that 'x is wrong' can express of an *activated* sentiment, i.e. a disposition to be angry/guilty at x plus anger/guilt at x. Or such a judgment can express an *unactivated* sentiment, i.e. a disposition to be angry/guilty at x only.

As mentioned earlier, moral judgments for Prinz also aim to assert a fact. He argues that "sentiments represent secondary qualities, but sentiments also dispose us to experience emotions, and emotions do not represent secondary qualities; they represent concerns... Thus, when an emotional state is generated from a sentiment, it represents both the secondary quality tracked by the sentiment and a concern" (Prinz 2007, 101). If this is correct, when my anger (emotion) at the pickpocket can be triggered by my disposition to feel anger at pickpockets (sentiment of disapprobation) and the act of pickpocketing itself, the anger represents the perceived property of *offensiveness* in the act of pickpocketing and the sentiment of disapprobation represents the perceived property of *wrongness* in the act of pickpocketing. Now this sets Prinz up to argue that moral judgments are truth apt. Given the representational features of moral sentiments, moral judgments that express moral sentiments are the sort of thing that can be true or false. The moral judgment that 'pickpocketing is wrong' (that expresses a sentiment) is true as long as someone has a sentiment of disapprobation toward pickpocketing. And the same moral judgment is false if no one has a sentiment of disapprobation toward pickpocketing. As Prinz argues, "[a]n action has the property of being morally wrong (right) just in case there is an observer who has a sentiment of disapprobation (approbation) toward it" (Prinz 2007, 92). Here we see another point of departure from the emotivist/expressivist position, i.e. for emotivists/expressivists moral judgments are *not* truth apt (at least not in the correspondence to fact manner, if at all), whereas for dispositionalists

like Prinz moral judgments are truth apt (in the correspondence to fact manner). And while there are differences between Prinz and Slote on *how* moral judgments are truth apt, they are united in the claim that moral judgments *are* truth apt (in the correspondence to fact manner).

With the above, we see Prinz's position on *verbalized* moral judgments. We are also now in a position to consider Prinz's view on moral judgments understood as a psychological state. I emphasize that I outline Prinz's theory on both moral judgments as utterances and as a psychological state as his *Emotional Construction of Morals* puts me in a unique position to do so. Prinz's work explicitly discusses these two types of moral judgments. This is unlike emotivists like Stevenson whose focus is largely on moral judgments as utterances/propositions. Prinz's work is also unlike Blackburn's *Ruling Passions* that hardly uses the term moral judgment at all. Gibbard's *Wise Choices, Apt Feelings* does appear to refer to moral judgments as utterances/propositions and as a psychological state of norm acceptance. However, Gibbard is not as explicit as Prinz in distinguishing between them. This point applies also to Slote's empathy-based sentimentalism. As such, to avoid doubt, by outlining Prinz's position on moral judgments as a psychological state, I am not suggesting in any way that the prior theorists discussed do not speak of moral judgment as a psychological state. With the caveat in mind, let us begin. Prinz argues that:

"moral concepts [as mental representations] dispose us to feel other-blame emotions when we are victims of transgressions and self-blame emotions when we transgress. If you believe that ϕ -ing is wrong, on this approach, you have a long-term memory representation that disposes you to feel guilt or shame if you ϕ , and anger, contempt, or disgust if someone else ϕ s. On any given occasion, when you have the occurrent thought that ϕ -ing is wrong, only one of these emotions will manifest itself... We can think of sentiment in long-term memory as standing belief, and the emotion in working memory as an occurrent belief. Or to introduce a useful piece of terminology, we can call the sentiment a *moral rule* and we can call a particular emotional manifestation of that sentiment a *moral judgment*." (Prinz 2007, 96).

The first point that I want to make clear here is that an agent's moral judgment (as a psychological state) that x is wrong can be her recognition that she has a disposition to feel 'other-blame' and 'self-blame' emotions when contemplating x or when x occurs. For such a judgment, the emotion need not be 'online'. The second point is that an agent's moral judgment (as a psychological state)

that x is wrong can be an emotional manifestation of a sentiment. If an agent has a sentiment of disapproval toward x and thinks of committing x nonetheless, that agent can, as a result of having such a sentiment, feel guilt for having thoughts of committing x . This guilt at thinking about committing x (e.g.) is also an instance of a moral judgment. Earlier in this section, I referred to a similar compound psychological state of anger at pickpocketing. Just as “[t]his compound constitutes the judgment that pickpocketing is wrong” (Prinz 2007, 96), the compound state of guilt at the thought of committing x can constitute the judgment that x is wrong. In sum, in terms of moral judgments as psychological states, there can be ‘cool’ dispassionate judgments (illustrated by the first point) and ‘hot’ passionate judgments (illustrated by the second point).

Pluralism

In the sections on Gibbard and Slote, we can see glimpses of how one can be a pluralist about moral judgments. In the preceding section on dispositionalism, we can see this even more clearly. One can argue that there are distinct types of moral judgments. Under Prinz’s theory, we can plausibly say that a moral judgment can be (e.g.) the utterance that ‘ x is wrong’ voicing a unactivated/activated sentiment of disapproval, the psychological state of having a sentiment of disapproval toward x , or the psychological state of feeling guilty/angry at x (as a manifestation of the sentiment of disapproval toward x). And with the distinct types of moral judgment involved, one can explain why emotions need not be ‘online’ whenever we make moral judgments. When we genuinely *utter* ‘ x is wrong’, we can be in an emotional state (where the sentiment is activated), but this need not be the case (where the sentiment is not activated). This applies also to the case where we genuinely *think* that x is wrong. When entertaining such a thought, we can also be in an emotional state (where the sentiment is activated), but this need not be so (where the sentiment is not activated). This plurality of moral judgments nicely segues us into Zagzebski’s theory. She is also a pluralist about moral judgments in that she argues that there are distinct types of moral judgments but “all of them ultimately rest on emotional experience” (Zagzebski 2004, 53). Now let us consider Zagzebski’s pluralistic account of moral judgment.

As we have seen in the previous chapter, for Zagzebski, a person experiencing an emotional state sees the intentional object of the emotion as coming under a thick affective concept. For example,

to experience pity/anger is to see x as coming under the thick affective concept of being pitiable/offensive. An emotion is a kind of perception in the sense that it picks up certain value properties in its intentional object. She argues that “[a]n emotion is a state of feeling a characteristic way about something seen *as* rude, as pitiful, as contemptible, and so on. So, the intentional object of an emotion is seen as falling under a thick concept” (Zagzebski 2003, 114). This emotional perception can also be expressed in a propositional form: “Emotions, like beliefs, can be expressed propositionally... when someone says “That is offensive,” we usually take that judgment to be an expression of an emotional state” (Zagzebski 2004, 74). And such expressions of emotions are also what Zagzebski call Level 1 judgments or “ground level moral judgments” (Zagzebski 2003, 108). There are two points that I would like to emphasize here. First, Zagzebski sees ground level moral judgments as utterances/propositions that express emotion. Second, the moral terms used in ground level moral judgments are far richer than what we discussed before. Thin terms like ‘good’, ‘bad’, ‘right’, and ‘wrong’ are explained in terms of other thicker properties like ‘offensive’, ‘pitiable’, ‘shameful’, etc. (as ground level moral judgements are expressions of emotions and emotions pick up such thick affective properties).

As discussed in a previous chapter, emotion for Zagzebski is a unitary state in that when we see x *as* pitiable (e.g.) we also feel the emotion of pity. Given that ground level moral judgments express the emotional state that one is in/one’s seeing of x as falling under a particular thick affective concept, the relevant emotion must be ‘online’ when one makes a ground level moral judgment. But there is another type of moral judgment where emotion need not be ‘online’ when one makes them. Such judgments are what Zagzebski calls “Level 2 Judgments” (Zagzebski 2003, 120). In explicating such Level 2 Judgments, she distinguishes between the psychological state of seeing x *as* pitiable (e.g.) from seeing *that* x is pitiable. This distinction allows us to see how we can genuinely utter that ‘ x is pitiable’ without feeling pity. Zagzebski argues that:

“I cannot see something as rude without feeling offended in the characteristic way that goes with rudeness, but I can see or judge *that* something is rude without feeling offended. I can do that in a way that roughly parallel the way in which I judge colors. I cannot see something as red without seeing red, i.e., without having a sensation of red, but I can see *that* something is red without a red sensation... Similarly, I cannot see something as pitiful without feeling pity, but I can see *that* it is pitiful without feeling pity. When I judge that

something is pitiful without feeling pity, as when I judge that something is red without having a red sensation, my judgment that the property applies is indirect" (Zagzebski 2003, 119).

Considering the above, I can judge that x is pitiful without feeling pity. What appears to follow from this is that level 2 judgments *cannot* be expressions of emotion. Since I do not feel pity when I make the level 2 judgment that x is pitiful, this judgment cannot be an expression of my pity (as it is not felt). Instead, level 2 judgments are expressions of a certain type of propositions. To utter ' x is pitiable' in level 2 judgments is to "express the proposition *that* the object is pitiful" (Zagzebski 2003, 120). Emotion as such is only indirectly involved in a level 2 judgment in that without the former (without first seeing something as pitiable (e.g.) via the emotional state of pity) one cannot genuinely understand the term 'pitiable' in the unemotional judgment that ' x is pitiable'. As Zagzebski argues, "[l]evel 2 judgments also utilize thick concepts, and... the agent cannot understand these concepts without having had the feelings that go with them when they are part of level 1 judgments" (Zagzebski 2003, 120).

Zagzebski's account is aimed at providing "an explanation of how central examples of moral judgment are both cognitive and intrinsically motivating, but which also includes an account of how the motivational aspect of the judgment can be detached from it" (Zagzebski 2003, 107-108). Her account intends to capture and reconcile two intuitions. First, moral judgments are truth apt. The judgment that ' x is wrong' can be true or false. Second, moral judgments are intrinsically motivating. To genuinely make a moral judgment that ' x is wrong' necessarily involves being motivated (even though defeasibly) toward avoiding x . With this aim in mind, we see how Zagzebski departs from the stance of emotivists like Stevenson. Both Zagzebski and Stevenson are committed to the claim that moral judgments are motivating. For Zagzebski in particular, moral judgments have motivational force in virtue of their ties with emotion. But unlike Zagzebski, moral judgments for Stevenson are not truth apt. We can also see how Zagzebski departs from the stance of expressivists like Blackburn. Though they are committed to the motivational aspect and truth aptness of moral judgments, Blackburn's commitment to the latter aspect is distinct from Zagzebski's. Unlike Blackburn, Zagzebski's account of moral judgment "is compatible with the view that moral judgment is primarily about some aspect of the world outside the mind of the judge, so the objectivity of moral judgment has not been precluded" (Zagzebski 2003, 108). In this

sense, Zagzebski's stance on the truth aptness of moral judgments is closer to that of Slote and Prinz than of Blackburn. For Zagzebski, Slote and Prinz, to say that "'x has a moral attribute' is true' is not (as Blackburn puts it) an affirmation that *x* has a moral attribute. Let us consider their stances specifically. For Slote, 'x has a moral attribute' is true if and only if *x* has the property of goodness/badness which causes or tends to cause empathic attitudes of warmth/chill. For Prinz, 'x has a moral attribute' is true if and only if the observer has a sentiment of approbation/disapprobation toward *x*. As for Zagzebski, 'x has a moral attribute' is true if and only if *x* falls under the relevant 'thick affective concept'. They differ from Blackburn who argues that to say that "'x has a moral attribute' is true' is to affirm that *x* has a moral attribute. And the key difference between Zagzebski, Slote and Prinz on one end and Blackburn on the other is that for the former claiming that "'x has a moral attribute' is true' is more than just stating 'x has a moral attribute', whereas for the latter claiming that "'x has a moral attribute' is true' is no different from stating 'x has a moral attribute'.

In the preceding paragraph, we see how in some respects Zagzebski's theory differs from other EAM theorists. Another point of difference that I would like to focus on is how Zagzebski's pluralistic account of moral judgments differs from Prinz's. Prinz's theory offers us an account where there are mainly three different types of moral judgment: moral judgment as an utterance that '*x* is wrong' (e.g.), moral judgment as a psychological state of feeling guilty/angry at *x*, and moral judgment as a psychological state of having a sentiment of disapproval toward *x*. The first refers to utterances of a sort, the second and third to psychological states of a sort. Zagzebski's theory appears to confine moral judgments to utterances of a certain sort. It is utterances like '*x* is pitiable' that can be either level 1 or level 2 judgments. However, Zagzebski is (almost) silent on whether the *psychological state* that a level 1 judgment expresses is a moral judgment of its own right. More specifically, the question now is whether (e.g.) the unitary state of feeling pity/seeing *x* as pitiable can be understood as moral judgment. I argue that for Zagzebski this unitary state of feeling pity and seeing of *x* as pitiable is *not* a moral judgment in its own right. While we can understand Prinz's compound state of 'anger at pickpocketing' as a moral judgment, we cannot understand the unitary state of feeling pity/seeing *x* as pitiable as a moral judgment in its own right. Allow me to explain.

As seen earlier, Prinz argues that ‘pickpocketing is wrong’ can be a verbalized expression of the state of anger at pickpocketing. In addition, he also argues that “these verbalizations get their meaning from the underlying emotional states” (Prinz 2007, 100). Similarly, we can say that, for Zagzebski, the meaning of level 1 judgments like ‘ x is pitiable’ is derived from the emotional states that underlie them. The state of pity and seeing x as pitiable, can be expressed as ‘ x is pitiable’. When this happens, ‘ x is pitiable’ (especially the term ‘pitiable’) gets its meaning from the emotional state of pity. In this sense, the emotional state has primacy over the expression of the emotional state. If we accept this primacy of state over expression and have given moral judgment ‘status’ to the latter, it seems that one might argue that it is not unreasonable to grant moral judgment status to the former on grounds of its primacy. But perhaps this move is too quick. While we may grant that the emotional state has primacy, it does not follow that the emotional state is the only constitutive component of a moral judgment. There can be other constitutive components. For Zagzebski, an agent can see x as pitiable but *not* judge that x is pitiable. She argues that “[a]dmiration [an emotion] need not include the judgment that the object of our admiration is admirable, but if we trust our emotion, we will be prepared to make that judgment. We tend not to make the judgment if we are skeptical of our emotion state, or if we wish to withhold judgment until we have been able to reflect or investigate the admired object further” (Zagzebski 2017, 34). From this we can plausibly say that, for Zagzebski, the emotional state itself is not sufficient for there to be a moral judgment (as a psychological state). Further, there must also be a mental act of assent. In other words, moral judgment (as a psychological state) has two constitutive components: the state of seeing x as coming under a thick affective property *plus* the mental act of accepting the proposition represented by the state as true.

Now if we grant that the state of seeing x as pitiable (plus assent) is a kind of moral judgment, we can also ask whether the unemotional moral *thought* of seeing that x is pitiable (plus assent) can also be a kind of moral judgment. I argue that we can as well. If the level 2 judgment ‘ x is pitiable’ is a moral judgment in its own right, it is not unreasonable to consider the state of seeing that x is pitiable (plus the mental act of accepting the proposition represented by this state as true) as a moral judgment as well. After all, the former is an expression of the latter. It is worth noting that for Zagzebski the state of seeing that x is pitiable lacks motivational force (Zagzebski 2003, 167). However, the state of seeing that x is pitiable need not be completely void of any motivational force to act in certain ways. Zagzebski argues that “[w]hen I judge that someone is lovable without

feeling love, my judgement that the property applies depends upon a prior experience of love... When I make a judgment without emotion..., it is likely that I must still be able to imagine seeing something as lovable..., but even imagining requires a faint copy of the feeling of love..." (Zagzebski 2003, 167-168). Her point here is that a level 2 judgment that *x* has a particular thick affective property is only understood by the agent if she had a prior level 1 judgment. I think this can in our context reveal another point. What I gather from Zagzebski's argument is that traces of feeling remain in our memory when we make level 2 judgments. If so, it is plausible that there can still be a slight motivational nudge (albeit likely to be phenomenologically imperceptible) to act in certain ways when we see that *x* is pitiable without noticeably feeling pity. As Zagzebski argues, level 2 judgments "typically retain a degree of motivational force inherited from our past moral experience" (Zagzebski 2003, 172). This perhaps is what allows Glen Pettigrove and Koji Tanaka to argue that a sage (who is no longer prone to feel anger) can still be motivated to respond in "ways characteristic of those who are angry". For example, a sage can still be "motivated to correct her students when they have transgressed... and defend those who are the victims of another's wrongdoing" (Pettigrove and Tanaka 2014, 284). Level 2 judgments can also explain situations where an agent experiences moral struggle. An agent who makes a level 2 judgment may have some but insufficient motivation to act on her judgment. She may even experience strong desires that run contrary to her level 2 judgment. This is a classic 'the spirit is willing, but the flesh is weak' situation. The point I would like to put forward is that Zagzebski's account of moral judgment with may get 'the best of both worlds'. Her account can explain how the distinct types of moral judgments can have some motivational force (albeit imperceptible at times) while maintaining the position that we can make moral judgments without being in an emotional state.

In sum, Zagzebski's pluralistic account of moral judgment set outs four types of moral judgment:

- Moral judgment (utterance) as expressions of emotion.
- Moral judgment (utterance) as expressions of a proposition ultimately based on emotion.
- Moral judgment (psychological state) as the emotional state of seeing *x* as having a thick affective property (plus assent).
- Moral judgment (psychological state) as the state of seeing *that x* has a thick affective property (plus assent).

Addressing Some Other Concerns

Let us take stock for a moment. In this chapter so far I have outlined: (a) some empirical support for EAMs, (b) how for EAMs emotion can be essentially involved in moral judgment, (c) how for EAMs the emotion involved in moral judgment differs, (d) how for some EAMs there are different types of moral judgment where emotion's involvement with these judgments pans out somewhat differently, (e) how some EAMs can address the concern that emotional states involved in moral judgment are incompatible with the phenomena of moral discourse, (f) how some EAMs can address the concern that moral judgments are not truth apt, and (g) how some EAMs can address the concern that there can be no dispassionate moral judgments.

As we can see, the outline of EAMs in the previous section is not only a survey of prominent EAMs (introducing readers to differing token EAM theories) but also a showcase of the resources available to EAMs to address some concerns about them. This is important to show that EAMs provide sufficiently plausible accounts of moral judgments before considering whether EAMs are well-suited to explain moral unanimity and diversity. In this section, I aim to focus on further showcasing the resources available to EAMs in addressing some general concerns. The concerns I would address in this section are:

- EAMs cannot explain how rational deliberation influence the making of moral judgments.
- EAMs cannot explain how moral judgments place demands on us. Specifically, (a) how they place demands on us regardless of whether we have the relevant emotional state and (b) how they appear to be issued from reason.
- EAMs imply that the absence of moral obligation on a person follows from the absence of a relevant emotion in that person.

Let us start with the first concern. We can elaborate this concern in the following manner: the essential involvement of emotion in moral judgment does not sit well with “the common observation that we often arrive at moral conclusions through a process of rational deliberation” (Prinz 2007, 123). As Prinz argues “[w]e try to alter each others’ moral judgments by presenting arguments. In private ethical reflection, we take special care to think about all of the relevant facts and weigh evidence. If emotions were driving morality, rational means of moral deliberation

would be poorly suited for arriving at moral judgments” (Prinz 2007, 123). Prinz raises this concern against his own account of moral judgment. However, this concern applies also to the token EAM theories discussed as all of them in one way or another see emotion as ‘driving’ moral judgment. And EAMs would suffer from an impediment if they cannot explain the phenomena of how rational deliberation influence the making of moral judgments.²⁴

The short response to the above concern is that many EAMs allow room for rational deliberation to influence the making of moral judgments. The question now is how such room is given. Prinz answers this question by arguing that:

“[w]e often need to use reason to demonstrate that an action falls under a moral category... Suppose I say that prayer in school is wrong. I might add a reason: it is wrong because it discriminates against atheists and members of minority religions. You might reply that it does not discriminate. Or you might reply that prohibitions against school prayer discriminate against members of majority religion, and hence discriminate against more people than if the prohibitions were lifted. This is a rational debate” (Prinz 2007, 124).

In this example, the abstract moral judgment of both sides (considering the reasons given) is ‘discrimination is wrong/offensive’ (an utterance/a thought) or ‘anger, disgust or guilt at discrimination/seeing discrimination as disgusting or offensive’ (a psychological state). The categorisation issue involved here is whether (not) prohibiting prayer in school counts as ‘discrimination’. In this respect, there is room for rational deliberation in influencing the making of moral judgments. We can argue about or reflect on this categorisation issue. Upon considering the arguments regarding this issue, I may (e.g.) come to the conclusion that not prohibiting prayer in school is discriminatory. This reasoning process can influence the making of moral judgments in the following way. If I am already disposed to react with anger (e.g.) toward discrimination, by putting ‘not prohibiting prayer in school’ under the category of discrimination via rational deliberation, I would also have a disposition to react with anger toward ‘not prohibiting prayer in school’.

²⁴ I will (in this chapter) only briefly address this concern. My aim at this juncture is to show that there is enough going on with EAMs to address this concern. However, I will address this concern with more detail in my response to the ‘adjudication problem’ outlined in the penultimate chapter of my thesis.

Variants of this 'room for deliberation' response can apply to other EAMs. I will use Zagzebski's account of moral judgment as an example.²⁵ She argues that:

"[p]ropositions of the form "X is pitiful", "Y is contemptible," "Z is fearsome," and so on can be true, and they are true just in case X is pitiful, Y is contemptible, Z is fearsome, and so on. They are not true *because* an agent feels pity, contempt, or fear... A person can be wrong in her construal of something as... contemptible... She might see something as contemptible that is not contemptible... And she can come to recognize her mistake. We believe we are recognizing mistakes when we change our perception of thick properties without any change in our grasp of the descriptive properties" (Zagzebski 2004, 77).

Without delving deep into Zagzebski's moral realist stance, it would suffice to note that for Zagzebski my judgment 'prohibiting prayer in school is contemptible' (e.g.) is not true just because I feel contempt at such a prohibition. There can as such be room for rational deliberation on whether 'prohibiting prayer in school is contemptible' is true. I can see prohibiting prayer in school as contemptible but come to recognize via reflection/argument that it is not contemptible. As a result, I can change my mind and no longer hold on to my initial moral judgment that 'prohibiting prayer in school is contemptible'. We can probe a bit deeper into how this reflection/argument can take place. Zagzebski argues that "[w]hat usually happens is that emotions are revised in roughly the way we revise our beliefs. There are feedback mechanism that change our emotions when the consequences of acting on them are not as expected" (Zagzebski 2004, 92). Here we see a way in which we can reflect/argue about our judgments. I can revise my judgment (e.g.) that 'John is pitiable' upon a subsequent discovery that John is a con artist. He understands what usually evokes pity, mimics them to earn an extra buck. Upon reflection, when I saw John as pitiable, he was not actually in a state of suffering but faking it. He duped me into giving him \$100. The consequence of acting on my judgment that 'John is pitiable' led me to lose \$100 to a con artist. Now with the understanding that John is a con artist that dupes people of their money by mimicking suffering, I may no longer see John as pitiable or see that John is pitiable. Instead, I may

²⁵ I note that I do not see the need to show how each EAM discussed above allows room for rational deliberation to influence the making of moral judgments. My task here is not to vindicate every or any EAM from the concern that they cannot explain the phenomena of how rational deliberation influence the making of moral judgments. Rather, my objective here is to showcase that EAMs in general are likely to have the means to deal with this concern. If so, by showing how one or some EAMs have the means to do so would suffice in meeting the objective.

see John as contemptible or see that John is contemptible. I could reject my initial judgment 'John is pitiable' and utter a revised judgment about John, i.e. 'John is contemptible'.

The second concern that one may have against EAMs is that they cannot explain how moral judgments place demands on us. As Prinz argues, "Kantians insist that [moral demands] are categorical. This is said to present a serious challenge for anyone who grounds morality in emotions... [Categorical imperatives] are said to govern us regardless of our inclinations, passions, and desires. That is what it means to say they are categorical... [T]hey are [also] said to be commands of reason. That is the best explanation of how they could govern us independently of our passions..." (Prinz 2007, 128). Now as the reader can see, there are at least two sub-concerns here.²⁶ I will not be discussing the first sub-concern in detail here as I discussed this earlier in this chapter. Briefly, I note that contemporary EAMs have the resources to reconcile the emotional basis of moral judgments with a common intuition that we need not be in an emotional state to judge that *x* is wrong. For Prinz, a moral judgment expresses a sentiment. And a sentiment is a *disposition* to feel certain emotions. *X* being wrong is as such not dependent on whether a person is in any emotional state. In this sense, we can see that the moral judgment '*x* is wrong' 'governs' us regardless of whether we are in the relevant emotional state. This same point applies to Gibbard where to claim that '*x* is wrong' is to claim that it is apt to feel resentment/guilt at *x*. There is no need for resentment/guilt to be present for *x* to be judged as wrong. We have also seen how the making of Zagzebski's level 2 judgments does not require one to be in any emotional state.

Now as for the second sub-concern, as Prinz argues "[a]ccording to Kant, [categorical imperatives] are demands of reason. Reason requires us to avoid immoral conduct. Immoral conduct is irrational" (Prinz 2007, 130). One way of understanding Kant here is that reason provides sufficient grounds for the categorical imperative that '*x* is wrong' or '*x* should not be done'. This if correct threatens the key theme that runs across all EAMs, i.e. some moral judgments have an emotional basis. This is because Kant offers an alternative non-emotional basis (if not *the* basis) for moral judgments. Now it would take an entire thesis by itself to mount a plausible challenge against Kant's position. This is not what I intend to do. Instead, I would draw the reader's attention to one

²⁶ More precisely, Prinz highlights three sub-concerns. However, I chose to focus on two sub-concerns for the purposes of highlighting concerns that are relevant to what I discussed earlier in this chapter.

of Prinz's argument against Kant that is aimed at reminding readers that "the Kantian program remains highly contentious" (Prinz 2007, 128). If the Kantian program remains highly contentious, the threat level it poses to the claim that all moral judgments have an emotional basis can be lowered. Prinz argues essentially that if an imperative that 'x is wrong' or 'x should not be done' is grounded on reason, it "cannot do justice to the main intuitions underlying the hypothesis that moral commands are categorical", i.e. "the intuition that moral requirements answer a fundamental normative question" (Prinz 2007, 133-134). When told 'thou shalt not murder, we would not usually ask the fundamental normative question 'why should I avoid murdering people'. We usually accept the normative force of the rule as a given. Now why does Prinz think that Kant fails to 'do justice' to this intuition? Prinz argues that "[a] purely rational rule cannot have motivational effect on us... 'Why obey the rules of morality?' asks the sensible knave. Kant's reply is that it would be irrational not to. To which the knave has a simple reply, 'Why be rational?' Or, more exactly, the knave might say, 'I recognize that it is inconsistent to value myself while treating others poorly, but I simply don't mind being inconsistent in this respect. After all, by favoring myself I have so much to gain!' Here, Kant must fall silent" (Prinz 2007, 134). In other words, a purely rational rule cannot do justice to the above intuition as such a rule leaves room for the 'fundamental normative question' to be asked and left unanswered.

Now we see with the above that if a rule is grounded on reason, it appears to run contrary to the intuition that we are usually closed to asking the fundamental normative question. This apparent inability of Kant's approach to capture the above intuition can be contrasted by EAMs' ability to do so. We are usually closed to asking the fundamental normative question like 'why should I avoid murdering people' as moral judgments like 'murder is wrong', being tied to emotion, motivate us to act accordingly. Prinz offers Hume's answer to the knave's questioning. Prinz argues that, for Hume, "... no person who is *constituted like us* can disobey morality without incurring an emotional cost. Ordinary people, thankfully, don't raise knavish questions... Hume's reply explains why moral rules seem to have power over us. He closes the gap between moral rule and action by arguing that, whenever we see a rule as moral, we are thereby moved to act on it. That's why rules seem categorical" (Prinz 2007, 134). To put this in the context of the EAMs discussed earlier in this chapter, since (a) moral judgments express emotion, have emotion as referents, or are constituted by emotion and (b) emotion has a motivational aspect, it appears that

EAMs have the resources to explain why we are usually closed to asking the fundamental normative question.²⁷

We can now move on to the third concern. Samuel Kerstein argues that for EAMs:

“... it is the person’s displeasing sentiments, one such as unease or shame, that form the basis of her obligation to acquire the character trait of being just to strangers, or at least to act in a way that a person with this trait would have... On this account the basis for an agent’s obligation to do something is a displeasing sentiment she has when, after taking the ‘common point of view,’ she contemplates her not doing it or, perhaps, her not possessing the character of someone who would do it... If an agent does not have this sentiment, then she has no obligation” (Kerstein 2006, 135).²⁸

Now allow me to put Kerstein’s argument in our context. If the reader permits me to set aside the point Kerstein made on the ‘common point of view’, we can rephrase the argument as follows: If the judgment that ‘*x* is wrong’ or ‘*x* should not be done’ is grounded on a particular emotion, then ‘*x* should not be done’ is only binding on those who have this particular emotion. Those who do not feel this particular emotion are exempt from complying with ‘*x* should not be done’. Now this argument (if successful) would take the normative force out of moral judgments if a person does not have the relevant emotion. Presumably, the judgment (e.g.) ‘genocide should not be committed’

²⁷ There will no doubt be Kantian replies to Prinz’s argument against Kant. Prinz, for example, anticipates the following reply: “the Kantian will contend that moral commands would have greater weight if they issued from reason”. And as moral judgments are judgments of passion, they “can be trumped by other affective states. For Kant, moral commands are reasons no matter what we desire” (Prinz 2007, 134). In response to this reply, Prinz makes three claims. First, a moral judgment (emotion- or reason-based) can be trumped by other affective states. The Kantian approach, as such, shares the same flaw as the emotion-based approach. Second, a moral judgment (emotion- or reason-based), even if trumped by other affective states in one instance, can trump other affective states in another instance. The Kantian approach, as such, shares the same resilience as the emotion-based approach. Third, while emotion-based moral judgments are necessarily motivating or based on an emotional state that is motivating, reason-based moral judgments can be void of any motivation. Reason-based moral judgments do not, as such, “directly bid for the attention of the will [as emotion-based judgments do]” (Prinz 2007, 135). It appears from this that emotion-based moral judgments carry more weight. Again, this is not a knockdown argument against Kant/Kantians. The point here is again to highlight (a) how the Kantian programme remains contentious and (b) how EAMs have the resources to close the gap between rule and action.

²⁸ This is only part of Kerstein’s argument. As Blackburn argues, “Kerstein concludes, for the sentimentalist, that there are people who lie under no obligation to universal justice. Hence, there are no categorical imperatives, for the categorical imperative embraces everyone” (Blackburn 2006, 145). As the reader shall see, my focus is on the criticism that the moral judgment that ‘*x* should not be done’ grounded on emotion does not give rise to an obligation to avoid *x* on the part of those who do not feel the relevant emotion. I would not be addressing the point claiming that there are no categorical imperatives as a result.

would not morally bind those who have no negative emotion whatsoever toward genocide. This is problematic as it runs contrary to a common intuition that even if one does not have any inclination to avoid a moral wrong (e.g.) she is still bound to avoid the moral wrong.

Blackburn offers a rather lengthy response to Kerstein's argument. However, I intend to focus only on Blackburn's argument that Kerstein's criticism conflates two different projects, i.e. the anatomist's and moralistic projects. The anatomist's project is "to give an accurate and complete account of the states of mind that gain expression in moral thinking". And the moralistic project is "to give an account of the 'sources' of our obligations" (Blackburn 2006, 146). Blackburn argues that "[i]f you confuse these two projects, you might end up saying that moral obligations 'stem from' or 'are based in' psychological states, and thence infer that in the absence of psychological states, the obligations disappear as well. The anatomical view is then supposed to lead to bad morals..." (Blackburn 2006, 146). This distinction between the anatomist's and moralistic project is a useful resource for EAMs to mount a defence against Kerstein's criticism. It allows EAMs to 'have the cake and eat it' so to speak. What I mean by this is that it allows EAMs to claim that, on one hand, emotion is essentially involved in the moral judgment that 'x should not be done' and on the other, this essential involvement does not imply that those without the relevant emotion do not have the moral obligation to avoid x.²⁹ Now let us consider as (an example) Blackburn's argument that moral obligations stem from something other than our emotion:

"... your obligations as a parent stem from the dependency of your children, their needs, and the absence of other social resources provide a substitute if you fail to meet those needs. If you don't care about your child, you are in breach of the obligation that the child's need places on you. The obligation does not come and go according to your affections, any more than your debt comes and goes depending on whether you care about it... The obligations

²⁹ This is not to say that all EAMs discussed will find no difficulty in utilising this resource. Zagzebski's stance (e.g.) that "[o]bligation *comes out of* emotion and the perception of thick affective properties [emphasis added]" and that "the sense of obligation is an emotion directed toward my own potential acts in response to some situation" (Zagzebski 2004, 146) may face a difficulty in utilising this resource as she seems to tie obligation closely to emotion. However, this does not mean that the difficulty cannot be overcome nor that alternative resources are not available to be utilised to address Kerstein's criticism. And given that the aim of this section is to showcase how EAMs in general can address some concerns and I showcased how EAMs in general can address Kerstein's criticism, I do not intend to explore how Zagzebski's EAM in particular can overcome this criticism.

you lie under, like the debts you owe, don't decrease or disappear when you stop caring about them" (Blackburn 2006, 146).

In short, for Blackburn the source of moral obligation is distinct from the source of moral judgment. The source of moral obligation, as it appears, stems from non-emotional factors (the needs of one's children, the debt owed, etc.). If this argument is correct, then Kerstein's criticism above would at the very least remain contentious.³⁰

With the three concerns outlined and EAMs responses put forward in this section, I hope that I have showcased how EAMs have the resources available to address the abovementioned concerns. By this I do not mean that these responses definitively address these concerns (as it would be impossible for me to do so taking into account the chief aim of my thesis). Instead, this section has a more modest aim, i.e. showing that EAMs *have the potential* to address these concerns. By showing such potential, I have done my part in showing that there is enough 'going on' with EAMs that would make the question 'how would EAMs explain moral unanimity and diversity' interesting.

³⁰ Kerstein and/or other Kantians may reply with the following argument: while distinguishing between the anatomist's and moralistic projects allows EAMs to have the cake and eat it, this move comes at the cost of conceding too much room to non-emotional factors. More specifically, the move leaves EAMs with a theory that resorts to non-emotional factors to justify the judgment 'x should not be done'. As a rejoinder, one can argue that while Blackburn distinguishes the anatomist's project from the moralistic project to avoid a situation where obligation disappears when feeling disappears, other EAMs do not have to make this move. As seen earlier in this chapter, Zagzebski (for example) can argue that while the feelings associated with an agent's expression of the proposition that 'x is wrong' may be phenomenologically imperceptible, this does not mean that such feelings are absent when the proposition is expressed. Traces of feeling remain in our memory when such level 2 judgments are made. It is plausible then that the agent can still be slightly motivated to avoid x. If so, the obligation to avoid x never really disappears. EAMs, as such, appear to have the resources available to address Kerstein's criticism. Again, my point here is not that EAMs would win this debate over Kerstein, but to show that his criticism at the very least remains contentious.

Chapter Three

“... Mencius said, ‘One’s natural tendencies enable one to do good; this is what I mean by human nature being good. When one does what is not good, it is not the fault of one’s native capacities. The mind of pity and commiseration is possessed by all human beings; the mind of shame and dislike is possessed by all human beings; the mind of respectfulness and reverence is possessed by all human beings; and the mind that knows right and wrong is possessed by all human beings. The mind of pity and commiseration is humaneness; the mind of shame and dislike is rightness; the mind of respectfulness and reverence is propriety; and the mind that knows right and wrong is wisdom...” (Mencius c 300 BC: 2009, 6A6).

Mencius, But Why?

With emotion theories and EAMs set out in the preceding chapters, I can now engage in a comparative exercise with a Mencian account of moral judgment. Such a comparison is apt in general and considering the aim of my thesis. First, it would be apt in the sense that a venture into a Chinese EAM would provide a somewhat culturally unique perspective and enrich our understanding of EAMs in general. As Slote argues “Chinese thought has fundamental lessons to teach the West” (Slote 2015, 11). Second, as discussed earlier, my ultimate aim is to show how EAMs can explain moral unanimity and diversity. This would especially involve a survey of non-Western moral judgments. Considering this, it would also be apt to consider a non-Western EAM as a precursor to such a survey. In incorporating an alternative non-Western standpoint, EAMs can have an enhanced credibility in its explanation of moral unanimity and diversity. EAMs can enhance their credibility in the sense that the tools used to explain moral unanimity and diversity across cultures are not confined to a narrow scope of Western theories but also extend to an Eastern theory. By outlining a Mencian EAM alongside Western EAMs provides a broader epistemic base for EAMs to explain moral unanimity and diversity. It is on these premises that I outline Mencius’ EAM.

In this chapter, I will argue for a Mencian account of moral judgment whereby moral judgment as a psychological state can be constituted by the cognitive component of emotion. And a moral judgment that is not constituted by the cognitive component of an emotion is ultimately grounded on a moral judgment that is so constituted. A novelty of this chapter lies in the production of a new argument against Myeong-seok Kim's objection that may threaten the characterization of the Mencian account of moral judgment as an EAM. While I will address his objection in more detail later in this chapter, it would suffice to note at this stage that it involves the claim that one of the four Mencian 'sprouts' of moral judgment is non-emotive. If so, and if this non-emotive sprout does not work in tandem with the other three emotive sprouts, it may negatively affect the claim that the Mencian account of moral judgment is an EAM. My argument against his objection essentially challenges his interpretation of the *Mencius* that led him to the claim that one of the four sprouts is non-emotive.³¹ Another novelty lies in the manner I resolve the objection raised by Emily McRae and close the gaps in David Wong's position that put into doubt Mencius' commitment to what is known as 'non-baseline' moral judgments. I will detail her objection and his position later in this chapter, but it would suffice to note at this point that (without addressing them) they threaten the Mencian account of moral judgment that I put forward.³² The third novelty lies in my argument on how the Mencian account of moral judgment is uniquely apt to explain moral unanimity and diversity. I do so with reference to a modified version of Mencius' water and sprout metaphors. Specifically, I argue that while Mencius himself does not use them for the undermentioned purposes, the metaphors put the Mencian account in a unique position to plausibly explain (at least to an extent) why we panculturally share certain moral judgments and why our moral judgments so often diverge.³³

Mapping on to Western Emotion Theories

Before outlining the Mencian account of moral judgment, there is one preliminary point that I would like to address. In this section, I highlight a difficulty in categorizing the Mencian account as an *emotion-based* account of moral judgment. This difficulty lies in the absence of the use of the term 'emotion' in the *Mencius*.³⁴ As Franklin Perkins argues, "[t]he text, *Mencius*, contains no words

³¹ I detail Kim's objection and my response in the section titled '*An alternative source for baseline judgments?*'.

³² I detail McRae's objection and my response in the section titled '*Mencius' commitment to non-baseline judgments?*'.

³³ I address this in detail in the section titled '*The water and sprout metaphors?*'.

³⁴ The *Mencius* records the philosophical teachings of Mencius, a Chinese philosopher who lived during the Warring States period (403-221 BC) (*Mencius* c 300 BC: 2009, ix).

that correspond to... 'emotion'" (Perkins 2002, 207). The difficulty compounds, especially if we want to compare the Mencian account with the Western EAMs discussed. We risk comparing oranges with apples. One way to resolve this difficulty is to examine the psychological states referred to in the *Mencius* and identify whether these states 'map on' to what Western EAMs would (broadly) regard as emotions. If the answer is 'yes', we can plausibly rely on references to these psychological states in the *Mencius* to present an emotion-based account of moral judgment and engage in a comparative exercise with Western EAMs. Perkins argues that "... although [Mencius] never uses a word corresponding to 'emotion', Mencius speaks of particular states that we would consider emotions, such as commiseration, love and anxiety" (Perkins 2002, 207). The referent for 'we' here is likely to be Western views of emotion (Perkins 2002, 208). But Perkins does not really specify what these Western views are and how they are like the psychological states referred to in the *Mencius*.³⁵

In this section, I aim to show how two psychological states referred to in the *Mencius*, compassion and shame, are similar to conceptions of emotion in some Western EAMs. I will not be addressing whether a Mencian emotion theory is a thought-based, feelings-based or hybrid approach. This means that I would not be addressing what emotions *essentially* involve. Instead, I would provide a brief sketch on what emotions like compassion and shame (for Mencius) *usually* involve. I do this to assure the reader that despite the non-use of the term 'emotion' in the *Mencius*, it is still plausible to consider a Mencian account of moral judgment as emotion-based.

Let us start with compassion. Mencius argues that:

"[a]ll human beings have a mind that cannot bear to see the suffering of others... Now, if anyone were suddenly to see a child about to fall into a well, his mind would be filled with alarm, distress, pity, and compassion. That he would react accordingly is not because he would hope to use the opportunity to ingratiate himself with the child's parents, nor because he would seek commendation from neighbors and friends, nor because he would

³⁵ Perkins need not have done so given that this is not the aim of his work *Mencius, Emotion and Autonomy*. Nonetheless, there is a small gap for me to fill. And with Western emotion theories outlined in chapter one, I am in a unique position to do so.

hate the adverse reputation [that could come from not reacting accordingly]" (Mencius c 300 BC: 2009, 2A6).

David Wong in interpreting this passage addresses the question of what compassion involves. He argues that "[o]ne may feel alarm or distress at the sight of a child's falling into a well, but be physically unable to save the child. What remains in such cases is the perception of someone in suffering or in danger, the phenomenal quality of alarm and distress and its accompanying physiological changes, and an impulse to act that remains unconnected to action" (Wong 1991, 32). At this point we see that a standard episode of compassion appears to involve perception, feelings, and a motivational tendency.³⁶ We also see such an episode in passage 1A7 of the *Mencius*. In response to King Xuan's question on whether he can protect his people, Mencius reminds the king of his compassion:

"I have heard Hu He say while the king was seated in the upper part of the hall someone led an ox past the hall below [in the courtyard]. On seeing this, the king asked where the ox was going and was told that it was being taken to serve as a blood sacrifice in the consecration of a bell. The king said, 'Spare it. I cannot bear its trembling, like one who, though blameless, is being led to the execution ground.' ... With such a mind one has what it takes to be a true king... the king could not bear to see its suffering" (Mencius c 300 BC: 2009, 1A7).

Here the king's compassion appears to involve a perception of the ox as suffering, feelings of distress, and the motivation to spare it from slaughter.

Now what about shame? Mencius argues that "one who lacks a mind that feels shame... would not be human" (Mencius c 300 BC: 2009, 2A6). One question that can arise here is what for Mencius

³⁶ Mencius a la Wong's understanding of compassion goes beyond this point. Wong goes on to argue that there is a cognitive dimension involved in compassion. By using the same passage from the *Mencius* above, he argues that for Mencius "[t]he reaction is to feel alarm and distress... purely from *recognition* of the child's endangerment [emphasis added]" (Wong 1991, 32). However, I do not intend to probe further as the task here is not to determine whether Mencius a la Wong's 'compassion' falls under the thought-based or feelings-based approaches discussed in chapter one, but to show how roughly speaking does this conception of compassion map on to Western views. It is not the purpose of this section to set out Mencius' emotion theory. The purpose is a more modest one, i.e. to assure readers that East and West are not talking past each other when they talk about emotions.

is this psychological state of shame about. We can find a clue to the answer in the following passage from the *Mencius*:

“Suppose there are a basketful of rice and a bowlful of soup. If I get them, I may remain alive; if I do not get them, I may well die. If they are offered contemptuously, a wayfarer will decline to accept them; if they are offered after having being trampled upon, a beggar will not demean himself by taking them. And yet when it comes to a stipend of ten thousand *zhong*, I accept them without regard for decorum and rightness. What do the ten thousand *zhong* add to me? Is it because I can then get beautiful dwellings that I take them, or the service of wives and concubines, or the gratitude of poor acquaintances that I share them with? What formerly I would not accept even at the risk of death, I now accept for the service of wives and concubines” (Mencius c 300 BC: 2009, 6A10).

Now for Mencius the beggar would not take food trampled upon as she sees such an act as shameful. Mencius considers the desire to avoid shame strong enough for one in this instance to risk even death. He also appears to imply that there are expectations on a civil servant’s or government official’s receipt of stipends from rulers. They must be received with ‘decorum’ and ‘rightness’. If the ruler is corrupt, receiving stipends from him is akin to receiving food that is trampled upon. The civil servant or government is expected to refrain from receiving such stipends. After which, Mencius questions why such civil servants or government officials would not see accepting big fat paychecks from corrupt rulers as shameful and desire to avoid it. The argument that Mencius appears to be putting forward here is essentially that if we see receiving food trampled upon as shameful, we should also see accepting such paychecks as shameful. And if we see accepting such paychecks as shameful, we would avoid accepting them. Now if my explication is correct, we see that a standard episode of shame involves the perception of certain acts as shameful, feelings of shame, and the motivational tendency to avoid such shameful acts.

With brief accounts of Mencian compassion and shame outlined, let us also briefly consider whether they ‘map on’ to what the West regards as standard episodes of emotions. As discussed in chapter one and I reiterate it here, “[s]tandard episodes of emotions involve a number of distinct components, such as what is often called a ‘cognitive basis’ – you need to perceive the drop of the plan, or to see your friend, in order to experience any emotional reaction – as well as a kind of possibly unarticulated appraisal, physiological changes, facial expressions, characteristic feelings,

cognitive and attentional processes, and finally, a motivational component” (Tappolet 2018, 495). As with Tappolet’s description, we have seen how Mencian compassion and shame can involve perceptions of a certain sort, characteristic feelings and motivational tendencies. In light of this, Mencian compassion and shame appear to correspond to what the West in general would regard as emotions. From the face of it, compassion and shame ‘map on’ well.

We can also consider how Mencian compassion and shame correspond to their Western counterparts. Given that Roberts (as discussed in chapter one) uses the defining proposition methodology to individuate emotions, we can conveniently compare Mencius’ compassion and shame with Roberts’ defining propositions for these emotions. For Roberts, the defining proposition of compassion is: *“It is important for S to be flourishing, but S is in distress or deficient in X way; may S’s distress be relieved or S’s deficiency made up”* (Roberts 2003, 295). As we can see there is a clear affinity between Roberts’ ‘Western’ compassion with the Mencian account. Just as Roberts’ compassion involves the construal of S as being in distress, Mencian compassion involves the perception of the ox as being in distress. And just as Roberts’ compassion involve the motivational tendency to relieve S’s distress, Mencian compassion involves the motivational tendency to spare the ox from slaughter. Now what about shame? For Roberts, the defining proposition of shame is: *“I am or appear unrespectable (unworthy, disgraced) in some way that it is very important to be or appear respectable (worthy); may I be or appear more respectable”* (Roberts 2003, 230). Again, we can see the affinity between Western and Eastern conceptions of shame. Just as Roberts’ shame involves the construal that I am disgraced in some way, Mencian shame involves the perception of receiving food trampled upon as disgraceful. And just as Roberts’ shame involve the motivational tendency to avoid what is construed as disgraceful, Mencian shame involves the motivational tendency to avoid receiving food trampled upon.

With the above, I hope I have assured the reader that the non-use of the term ‘emotion’ in the *Mencius* does not hamper the plausibility of a Mencian *emotion*-based account of moral judgment and that the risk of comparing the ‘apples’ of Western emotions with ‘oranges’ of Eastern emotions does not appear to be a high one. Though Mencius does not use the term ‘emotion’ per se, he appears to refer to psychological states that ‘map on’ to Western views on emotions. Now that I have addressed this, we can consider the role Mencian emotions play in morality.

From Virtues to Judgments

For Mencius, certain emotions play a central role in morality. He argues that: “The mind’s [xin 心] feeling of pity and compassion is the sprout of humaneness [ren 仁]; the mind’s feeling of shame and aversion is the sprout of rightness [yi 義]; the mind’s feeling of modesty and compliance is the sprout of propriety [li 禮]; and the mind’s sense of right and wrong is the sprout of wisdom [zhi 智].” (Mencius c 300 BC: 2009, 2A6). With this we see that for Mencius certain emotions (pity/compassion, shame/aversion, modesty/compliance, and a sense of right and wrong) are the origins of the Confucian virtues of *ren*, *yi*, *li* and *zhi* (Virág 2016, Paragraph 9).³⁷ These four sprouts are also commonly understood as “natural emotion[s]” (Perkins 2002, 219). This gives the reader a sneak preview of why Mencius’ moral philosophy can be helpful in showing how EAMs can explain moral unanimity and diversity. For starters, Mencius argues that there is an essential connection between emotions and moral *virtues*. This claim is close to an EAM claim, i.e. there is an essential connection between emotions and moral *judgments*.

The sprout metaphor in 2A6 of the *Mencius* specifies a set of emotions that ground the Confucian virtues of *ren*, *yi*, *li* and *zhi*. And as Owen Flanagan, commenting on 2A6 of the *Mencius*, argues, “[t]he core idea is that what we call character or conceive of as moral personality, even as genuinely unified experientially, or as an interactive set of virtues, norms, and rules, even as a unity of virtues, is, to some significant degree, the emergent product of growth, regulation, moderation, modification, tuning, and suppression of ancient equipment composed of relatively narrow dedicated processors [i.e. the four sprouts]” (Flanagan 2016, 57). From this we can begin to see how for Mencius these emotions are essential for the possession of *moral virtues*.

Mencius’ grounding of the virtues in the emotions is set against the backdrop of rival theories present in his time:

³⁷ Emotions here are understood broadly. Instead of confining emotions to emotions strictly so called, this understanding includes that which is emotive or affective under the banner of emotions.

“The words of Yang Zhu and Mo Di flow throughout the world; the teachings circulating in the world today all go back to Yang or Mo. Yang holds for egoism, which involves denial of one’s sovereign; Mo holds for impartial care, which entails denial of one’s parents... If the ways of Yang and Mo are not stopped, and the way of Confucius is not made known, the people will be deceived by these deviant views, and the path of humaneness and rightness will be blocked” (Mencius c 300 BC: 2009, 3B9).

And as Bryan Van Norden argues, “Yang Zhu and Mozi represent two extreme positions that Mengzi rejects. Yang Zhu held that benevolence and righteousness are unnatural, the products of social condition. A human who follows his nature will act only in his own self-interest. Thus, in modern terms, Yang Zhu may be classified as an ‘ethical egoist.’ Mozi was at the opposite extreme, advocating impartial concern for everyone. He thus rejected ‘differentiated love,’ as well as the ritual practices that Confucians saw as cultivating virtue” (Mencius c 300 BC: 2008, 85). With the above, we see that for Mencius Yang’s excessive focus on one’s own welfare is at the expense of inherent human tendencies that are other-directed. And Mo’s excessive focus on impartial care that does not distinguish between duties toward kin and non-kin is alien to the inherent human tendency to care first for one’s kin. Without going too much into the theories of Yang and Mo, it would suffice to say that for Mencius both theorists propose an ‘unnatural’ grounding of morality. Contrary to Yang and Mo, Mencius’ stance is that morality is grounded in certain inherent human tendencies. And it is necessary for one to have certain emotional dispositions to possess these moral *virtues*. As the reader shall see later in this chapter and the next, Mencius’ claim that we have certain inherent emotional tendencies can help explain moral unanimity and diversity. We will return to this point later; while we have firmly established the nexus between emotion and moral *virtue*, we are yet to outline the nexus between emotion and moral *judgment*. It is to this latter point that I now turn.

Mencian Pluralism

For the Mencian account I am proposing, moral judgments ultimately rest on emotional experience. This claim requires at least two further clarifications: (a) what counts as Mencian moral judgments, and (b) how do such moral judgments ‘rest on’ emotional experience. I answer these questions by using Wong’s interpretation of the *Mencius* in *Moral Sentimentalism in Early Confucian Thought* (2017) as a starting point. My reasons for starting with Wong are twofold. The first reason is that Wong’s

work is in line with the overall aim of this chapter. He explicitly engages in a comparative exercise between Western moral sentimentalism and Mencian ethics. Second, there is room for further exploration given Wong's cursory treatment of Mencius' account of moral judgments. I intend to clothe the 'bare bones' of Wong's stance, with 'flesh' from various sources (including the *Mencius* itself, secondary literature on the *Mencius* and Western philosophical literature). I emphasize that Wong need not endorse this clothing with flesh. I start with Wong but would not end with him. I do not aim to present Mencius a la Wong. Rather, I aim to utilise Wong's stance as an 'overture' to my take on a Mencian account of moral judgments. Let us begin.

Mencius a la Wong is a pluralist about moral judgments. For him, there are at least two types of moral judgments. They are baseline judgments and non-baseline judgments. Let us start with the former. Wong argues that for Mencius we are predisposed "to feel and act in morally appropriate ways and *to make intuitive normative judgments* that can with the right nurturing conditions give human beings guidance as to the proper emphasis to be given to the desires of the senses [emphasis added]" (Wong 2018). He also argues that "on one plausible interpretation of Mencius, Heaven endows human beings with the dispositions to make correct intuitive judgments, e.g., that someone is suffering and in need of help, or that accepting food offered with abuse or in an insulting way would be a shameful thing to do. Such judgments are made spontaneously and without deliberation or calculation, and are rooted in the inborn dispositions that form the beginnings of morality" (Wong 2017, 240). Now as we can see, Mencius a la Wong characterizes a baseline judgment as spontaneous, non-deliberative, and rooted in certain inborn dispositions. Wong however does not explicitly address whether the baseline judgment is a psychological state of judging that x has a moral attribute, the utterance that ' x has a moral attribute' or both. On this point, I argue that it is plausible to understand a baseline judgment as a psychological state of judging that x has a moral attribute and restrict our discussion on intuitive judgment to this. The understanding I propose is plausible as baseline judgments share very similar features with Daniel Kahneman's "System 1" mode of thinking of which "operates automatically and quickly, with little or no effort and no sense of voluntary control" and "generates impressions, feelings, and inclinations..." (Kahneman 2011: 2012, 20, 105). What is clear here is that System 1 is a mode of *thinking* (a psychological state). And there are sufficiently similar characteristics between baseline judgments and System 1, i.e. automatic, quick, non-deliberative, etc., to suggest that the former is also a psychological state. When I encounter a child about to fall in the well, I may react

automatically and without deliberation with the baseline judgment that the child's life is endangered and needs help. Or when someone attempts to give me food trampled upon, I may react automatically and without deliberation with a baseline judgment that accepting it is shameful.³⁸

Now what about non-baseline judgments? Wong argues that "'baseline' judgments do not provide the answers for all occasions on which one asks what is the right thing to do... However, they form a kind of basis for deriving the answers through analogical reasoning. When we are wondering what to do, we try to find the closest analogy between the present situation and a past situation in which we have made a baseline judgment. We then identify a response to the present situation that most resembles the response embodied in the baseline judgment" (Wong 2017, 240). In short, non-baseline judgments are judgments derived from baseline judgments. Wong provides an example of how this derivation takes place: "a baseline judgment that has to do with compassion is that it is right or appropriate to try to prevent a child from falling into a well (*Mencius* 2A6). A feature of that situation that Mencius holds is relevant to its rightness is the child's innocence (3A5). Reflecting on whether one should help another person, therefore, involves reflecting on cases such as the child and the well, and then determining whether the person in question is suffering and innocent" (Wong 2017, 240).³⁹ Here we see that the relevant features leading to the baseline judgment are the 'suffering' and 'innocence' of the child. These features can assist the agent in determining what should be done in another situation. If a prisoner exhibits clear signs of distress when led to the gallows, and the agent believes that the prisoner was wrongly convicted, she may make the non-baseline judgment that 'it is right to help the prisoner'.

³⁸ To avoid doubt, when I say that such baseline judgments are a psychological state, I do not mean that these judgments necessarily occur in our minds as explicit propositions. Instead, I graft Roberts' approach to emotional states (discussed in chapter one) to baseline judgments. Just as emotional states can be propositional in the sense that "they are susceptible of reasonably accurate propositional characterization" (Roberts 2009, 573), intuitive judgments need not be explicitly propositional when they occur and but we can subsequently characterize them in propositional form.

³⁹ A problem with Wong's outline of non-baseline judgments here is that it appears speculative. The textual evidence used to support this outline appears 'weak'. Neither 2A6 nor 3A5 explicitly describe the process of deriving non-baseline judgments from baseline judgments. I will address this point further later in this chapter.

Emotive Source of Moral Judgments

In the above section, we see that there are baseline judgments and non-baseline judgments. We also see that the latter is derived from the former. In this section, we can see how Mencian moral judgments have an emotive source. In 3A5, Mencius argues that morality has only 'one root' and that root is the emotions. It is to this passage that I turn to as textual evidence to support a Mencian EAM:

"Yi Zhi, a Mohist, sought, through Xu Bi, to meet Mencius. Mencius said, 'I certainly want to see him, but now I am still sick. When I have recovered from my illness, I will go myself to see him. Master Yi need not come here.'

Another day Yi Zhi sought again to see Mencius. Mencius said, 'Now I am able to see him. If I do not correct him, the Way will not be made apparent. I will correct him. I have heard that Master Yi is a Mohist. In regulating funeral practices, the Mohist way is that of simplicity, and Master Yi is contemplating changing the world accordingly. What makes him think that, unless the deceased are buried this way, they are not honored? Master Yi himself buried his parents in a lavish style, thus serving his parents in a way that he himself disparages.'

Master Xu informed Yi Zhi, who said, 'According to the Confucian way, the ancients acted as if they were protecting an infant. What does this teaching mean? To me it means that one should love without distinctions but that the love begins with parents and extends from there.'

Master Xu told this to Mencius. Mencius replied, 'Does Master Yi believe that a man's affection for his brother's child is just like his affection for the child of a neighbor? What he should have taken from the teaching [he cited] is that, if a child crawling toward a well is about to fall in, this is not the fault of the child. Heaven, in giving birth to living beings, causes them to have one root, while Master Yi supposes they have two roots'" (Mencius c 300 BC: 2009, 3A5).

As discussed earlier, Mencius rejects Mohist thinking as it proposes an 'unnatural' grounding of morality. In the above passage, we now see how this rejection pans out. Essentially, we see here a disagreement between Yi, a Mohist, and Mencius on how a Confucian teaching should be understood. Mencius argues that Yi is proposing a 'two root' understanding, whereas the correct

understanding is that of 'one root'. And as David Nivison argues, at this point in the passage, "[w]hat Mencius means by 'one root' and 'two roots' is not immediately evident" (Nivison 1980, 741). A further explanation on Mencius' part is needed to understand this 'one root'/'two root' dichotomy. Mencius then continues with a narrative:

"Now, in high antiquity there were some who did not bury their parents. When their parents died, they picked them up and cast them into a ditch. Another day, when they passed by, they saw that they were being devoured by foxes and wildcats and bitten by flies and gnats. Sweat broke out on their foreheads, and they averted their eyes to avoid the sight. The sweat was not because of what others would think but was an expression in their faces and eyes of what was present in their innermost hearts. They returned home and brought earth-carrying baskets and spades to cover them over. Burying them was truly right, and filial children and benevolent people also act properly when they bury their parents" (Mencius c 300 BC: 2009, 3A5).

With the above narrative, we can now start to 'piece together' the 'one root'/'two root' dichotomy. Here we see how a certain psychological state *motivates* the action of burying one's dead parents. Other Mencian scholars also share this interpretive point. Curie Virag, for example, argues that "[w]hat drove these negligent offspring to return and give their parents a proper burial was a certain irrepressible emotion" (Virag 2017, 117). And Nivison argues that we should understand Yi Zhi's position in the following manner:

"I have a *basic affection-capacity* which reveals itself in a basic way- in this case as a parental and familial affection. Having this capacity, I am then able to apply it to others, in accordance with my *beliefs* about how it should be focused – i.e., in accordance with the doctrines to which I adhere or my moral reasons for these doctrines. Morality on this view depends on two things, which are independent of each other: what I *think* I should do, and could state in *words* and reason about; and my capacity to feel certain emotions, which I can steer and shape so as to be *moved* to do what my principles tell me I should" (Nivison 1996, 102).

With both Yi Zhi and Mencius' positions set out, we see that Mencius rejects 'doctrines' as a source of moral motivation. Emily McRae endorses Nivison's argument by arguing that "[a]ccording to Mengzi [Mencius], the trouble with Yi Zhi is that he mistakenly believes that there are two sources that contribute to moral development – one's basic moral feelings and attitudes and one's

endorsement of a moral doctrine or philosophical system – and that they both have *equal motivational force* [emphasis added]” (McRae 2011, 591). As such, we can first understand the one root/two root dichotomy as a disagreement about the sources of moral motivation. Yi Zhi argues that emotion *and* doctrines are the sources of moral motivation. Mencius rejects this ‘two root’ claim. Instead, Mencius puts forward the claim that only emotion is the source of moral motivation, i.e. the ‘one root’ claim.

With the above, 3A5 reveals Mencius’ argument that emotion serves as the ‘one root’ for moral *motivation*. As a corollary, we can also say that 3A5 reveals Mencius’ argument that emotion serves as the ‘one root’ for moral *judgment*. Not only do we see that the children being in an emotional state were *motivated* to return home to get the relevant equipment to bury their parents, we also see in 3A5 that the children being in the emotional state *saw that* burying their parents ‘was truly right’. The children after seeing the decomposing bodies of their parents were in some distressed state. They could not bear seeing the decomposed bodies of their dead parents and the bodies being devoured by animals. It is via this unnamed emotional state that the children saw that burying their parents ‘was truly right’. As an aid to this interpretation of 3A5, I draw on Zagzebski’s argument that “[a]n emotion (motive) directed toward a potential act or omission can sometimes be experienced not only as strong but as leaving room for no other option” (Zagzebski 2004, 146). For Zagzebski, we can (via an emotional state) see that *x* ought to be or must be done. Similarly, for Mencius, we can not only be driven by the motivational force behind the irrepressible emotion to bury our dead parents, but we can also be driven by the same force to see that we must bury our parents. It may be this motivational force that led the children to see that they must bury their children and bring their spades and baskets from home to do so. This is a way in which we can plausibly understand 3A5 as putting forward the claim that emotion serves as the ‘one root’ for our source of moral judgment. Nivison’s argument regarding 3A5 is that for Mencius “what we really want to do [moral motivation], and *what we ought to do* [moral judgment] and can rightly recognize we ought to do, in the last analysis, both have their source in the human heart, the ‘one root.’ [emphasis added]” (Nivison 1980, 742).⁴⁰ The interpretation of 3A5 I presented above can be regarded as a more detailed development of Nivison’s argument. Not only do I accept that

⁴⁰ Nivison understands ‘heart’ as “the locus of moral emotions or feelings” (Kim 2018, 52).

motivation and judgment find their source in the human heart, I took a step further in illustrating *how* moral judgments can have an emotive source.

The Nexus Between Emotion and Baseline Judgment

With the preceding two sections, we see that there are two types of moral judgments (baseline and non-baseline) and these judgments have an emotive source. We can now consider in more detail how this nexus between emotion and each type of moral judgment pans out. Let us begin with baseline judgments. Wong argues that “[i]f the feelings of shame/dislike, for example, can involve the thought that something would be *shameful*, perhaps the [four] hearts [or emotions] involve certain intuitive [or baseline] judgments that in principle are separable from the affective components” (Wong 2017, 239). Here he appears to argue for two points: (a) the four emotive sprouts can have cognitive and affective components and (b) baseline judgments are constituted by the cognitive component of these emotions.

In explicating points (a) and (b), I start by briefly outlining the cognitive and affective components that can be involved in the 2A6 emotions. Let us begin with compassion. We saw earlier in this chapter Wong’s argument that for Mencius this emotion appears to be constituted by perception, feelings, and a motivational tendency. We can categorize these features either as the emotion’s affective or cognitive components.⁴¹ As it appears, the affective component comprises the feelings

⁴¹ Wong is not alone in arguing that the above emotions comprise affective and cognitive components. This view is, at least at a high level, also one that is espoused by other Mencian scholars. My point here is that this interpretation of the *Mencius* is plausible not based on Wong’s argument alone. Rather, the interpretation is plausible as it finds its basis on the consensus of Mencian scholars. In this instance, Wong’s argument is the vehicle through which I convey a shared interpretation of 2A6. Flanagan, for example, argues that “[e]ach sprout [including compassion] is an innate perceptual-affective-cognitive-conative disposition that possesses the potential, the natural trajectory, to grow into one of the four cardinal virtues” (Flanagan 2016, 62). The difference between Flanagan and Wong here is a matter of classification. For Flanagan, the 2A6 emotions involve perceptual, affective, cognitive and conative components. As for Wong, he lumps the perceptual and cognitive under the heading of ‘cognitive component’ and the affective and conative under the heading of ‘affective component’. Apart from Flanagan, Virag also argues that “[w]hen Mencius refers to the spontaneous prompts of moral feeling [in 2A6], these are not, strictly speaking, simply emotive, for they also involve judgments and assessments based on how things are in the world” (Virag 2017, 131). She argues so to make clear that on one hand, we should recognise “the activities of thinking and feeling as intertwined” and on the other that we should not collapse the conceptual distinctions between them (Virag 2017, 131). To support this claim, she argues that for Mencius “the *xin* [heart] is the site of the ‘sprouts’ of morally inclined feelings that all human beings presumably possessed. But these four sprouts include not only such feelings as ‘pity and compassion,’ and ‘shame and aversion,’ but also ‘courtesy and modesty’ and ‘a sense of right and wrong,’ which are clearly not simply ‘emotive in a narrow sense but also involve

of alarm and distress and the impulse to act in helping ways. As for compassion's cognitive component, Wong argues that for Mencius:

“[t]o respond to someone's suffering is to *perceive* him or her as suffering. Furthermore, Mencius in 1A7 and 3A5 links the feeling of compassion with *perceiving* the one who is suffering as innocent, or as underserving of the suffering. King Xuan in 1A7 is led to spare an ox being led to ritual slaughter because its terror reminds him of an innocent man being led to execution. In 3A5, Mencius links the compassion we feel for a child about to fall into a well with her not having done anything to deserve such a fate [emphasis added]” (Wong 2017, 233).

What is clear from this is that compassion's cognitive component can involve a range of perceptions. The perceptions range from the less cognitively laden (e.g. seeing that *x* is suffering) to the more cognitively laden (e.g. seeing that *x* is innocent and as such undeserving of suffering). In addition, Wong argues that compassion's cognitive component can involve another perception: “compassion also involves judging that one has a reason to help” (Wong 2002, 195).⁴² Wong continues to argue that ‘reason’ here is normative reason, it is “equivalent to ‘something that counts in favor of doing something.’ To cite someone's (one's own or another's) reason for doing something is to provide a justification for it or to explain why that person thought it was justified, or both” (Wong 2002, 195). Here we see the involvement of another layer of perception. Not only does compassion involve the perception of *x* as suffering and innocent, this emotional state can also involve the perception of *x* as suffering and innocent *as a reason* to alleviate *x*'s suffering.

It is important to note that Wong's setting out of compassion's cognitive and affective components (and the other 2A6 emotions below) does not involve an analysis of emotion concepts “in terms of necessary and sufficient conditions” (Wong 1991, 56). Wong is not arguing, for example, that the concept of compassion necessarily tracks the perception of *x* as suffering and innocent *as a reason* to alleviate *x*'s suffering. Instead, Wong's setting out of the cognitive and affective components of

a more complex range of perceptions, evaluations, and judgments” (Virag 2017, 110). What the above implies is that compassion (e.g.) do not just involve feelings, they also involve judgments about the world.

⁴² Though Wong uses ‘judgment’ in this instance, in other parts of his *Reasons and Analogical Reasoning in Mengzi*, he uses the term ‘perception’ instead (see Wong 2002, 194, 196). What he likely means by these terms is that to judge/perceive that the situation is a reason to act in certain ways is to *recognize* the situation as a reason to act in certain ways. As such, to avoid confusion and for consistency's sake, I would in this chapter refer to this judgment/perception/recognition as perception.

2A6 emotions is aimed at making clear the paradigmatic instances of such emotions. Allow me to explain this in the context of the debate between Wong and Craig Ihara. Ihara argues that “it is quite conceivable for someone to have a moral emotion such as compassion and, because of social conditioning, not regard it as identifying any justifying reason for action” (Ihara 1991, 47). Ihara’s argument here is that the emotion concept does not appear to track the cognitive state of recognizing justifying reasons. As Ihara argues, “emotions in Mencius are much more visceral than Wong believes. This is compatible with saying that they are cognitive in... that they help to sort out features of particular situations on the basis of salience, but this does deny that they entail the recognition of general justifying reasons for acting” (Ihara 1991, 52). Wong’s response to this objection is to refrain from analyzing the concept of compassion ‘in terms of necessary and sufficient conditions’. He argues that “[t]he concept of compassion, like many others, involves a typical or paradigmatic case. Not only is there a paradigm scenario that constitutes the intentional object of compassion, but there are other features of a *paradigmatic* case of compassion, such as helping behavior, and... implicit recognition of a reason to help. When all the features we typically associate with compassion are present, we have what we would call a fully determinate, unambiguous instance of compassion” (Wong 1991, 56). Here Wong’s response is that this perception of *x* as suffering and innocent as a reason to alleviate *x*’s suffering is a *paradigmatic (but not a necessary)* feature of the cognitive component of compassion. When we are in the emotional state of compassion, we can but need not have such a perception. Compassion can be more ‘visceral’ or less ‘full-blooded’ in the sense that it does not involve such a perception. With this understanding that Wong is not analyzing emotion concepts in terms of their necessary and sufficient components but only outlining the paradigmatic instances of such emotions, let us consider briefly the remaining 2A6 emotions.

In terms of shame, I argued earlier in this chapter that in the *Mencius* a standard episode of shame involves the perception of certain acts as shameful, feelings of shame, and the motivational tendency to avoid such shameful acts. The affective component of shame can as such involve the feelings of shame and the motivational tendency to avoid a shameful state. The perception of an act as shameful can then be the cognitive component of shame. This is similar to Wong’s argument that “shame... also include thoughts that it would be shameful to perform a certain action. 6A10 [see above] provides an example of this feeling in the beggar who rejects food offered with abuse on the grounds that it would be shameful to accept the food” (Wong 2017, 233). From this passage

and by applying Wong's schema of emotion involving the perception of seeing its intentional object as a reason for action, it appears that paradigmatic instances of shame involve another layer of perception, i.e. the perception that *x*'s shameful feature is a reason to avoid *x*-ing. Not only do I see receiving food trampled upon as shameful, I see also that the shameful receipt of food trampled upon as something to be avoided.

As for deferential feelings, Wong argues that "[d]eference and yielding are affective attitudes one adopts toward one's parents or elders, for example, and involve thoughts of what is proper and respectful" (Wong 2017, 233). Wong's argument here is terse. While the argument offers some help in outlining the affective component of deferential feelings, it does not offer much help in outlining their cognitive component. I intend to 'fill the gaps' with the resources available to me. Let us start with the affective component. The affective component of deferential feelings can involve the impulse to defer parents and/or elders and what is felt when experiencing this impulse. This claim is in line with Flanagan's argument that deference comprises the disposition to "experience an impulse to defer to humans or nonhuman animals who are larger or more powerful, which is eventually extended to those who are both older and wiser..." (Flanagan 2016, 57). Now what then of the cognitive component? We can plausibly imply from Wong's argument that the intentional object of deference is one's parents or elders. The perception involved in deferential feelings can be the seeing of one's parents or elders as worthy of respect. A paradigmatic instance of deference can also have an additional layer of perception whereby we see certain facts about one's parents/elders as a reason or reasons to act in respectful ways toward one's parents/elders. While this view cannot be read directly from the *Mencius*, it is a view that is inspired by Mencius' stand on the sprout of deferential feelings. Via deferential feelings, I may (for example) see the past sacrifices my mother has made to put food on the table for myself and my brother (facts about my mother) as a reason to give her first dibs on which restaurant we should go to for dinner (an act of respect).

As for the sense of right and wrong, Wong argues that "*shilfei*" [right and wrong] can have the connotation of accepting and rejecting, which is why they are sometimes translated as 'approval/disapproval.' So it is not definitive but plausible to interpret Mencius as conceiving all four of the hearts as simultaneously cognitive and affective. They can be conceived as affective

reactions that are infused with thought and meaning-conferring perception” (Wong 2017, 233). What we can see here is that for Wong *shil/fei* may also involve an affective and a cognitive component. However, he is again scant with an explanation on what these components involve. Flanagan may offer some help. He argues that *shil/fei* consists of the disposition to “experience certain things as right or wrong, fitting or not, fair or not...” (Flanagan 2016, 57). As for the cognitive-ladenness involved in *shil/fei*, it need not be a high one. As he also argues: “*zhi* [wisdom] is the mature form of an ability that begins in an innate tendency to respond with approval or disapproval to simple cases of good or bad character (sharers and non-sharers) or action (giving or withholding cookies). Very young children and perhaps other animals respond in these kinds of ways: approving – Yes! Yes! Smile, or disapproving – No! No! Scowl – of certain people and behavior” (Flanagan 2014, 83).⁴³ With Flanagan’s help, we can see that it is plausible to understand the cognitive component of *shil/fei* as involving the *perception* of certain people/behaviour as good/bad/right/wrong. As for *shil/fei*’s *affective* component, we can plausibly understand it as involving the feelings that accompany such perceptions. The smile of approval can be the facial expression of an internal state of liking the person/behaviour. And the scowl of disapproval, the expression of the internal state of disliking the person/behaviour.

In sum, a Mencian account of baseline judgments starts with the four emotive sprouts. Each sprout has cognitive and affective components. Their cognitive components are as follows:

- Compassion can involve perceptions of *x* as suffering, *x* as innocent, *x* as undeserving of suffering, and *x*’s suffering as a reason for helping *x*.
- Shame can involve perceptions of *x* as shameful act, and *x*’s shameful feature as a reason to avoid *x*-ing.
- Deferential feelings can involve perceptions of *x* (an action) as proper or respectful, and *x* (as a person/object) as worthy of respect, and facts about *x* (as a person/object worthy of respect) as a reason for acting in respectful ways.

⁴³ Flanagan’s position also enjoys some empirical support. An experiment conducted by Hamlin, Wynn and Bloom show that “6- and 10-month-old infants take into account an individual’s actions towards others in evaluating that individual as appealing or aversive: infant prefer an individual who helps another to one who hinders another, prefer a helping individual to a neutral individual, and prefer a neutral individual to a hindering individual” (Hamlin, Wynn and Bloom 2007, 557).

- The sense of right and wrong can involve perceptions of x (as an action) as right/wrong, and x (as a person) as good/bad.

These perceptions or cognitive aspects of the four emotive sprouts are constitutive of Mencian baseline judgments. This is how baseline judgments have emotions as its source and how the nexus between baseline judgments and emotions pan out.

An Alternative Source for Baseline Judgments?

Let us now consider one objection against the claim that baseline judgments have emotions as its source. Myeong-seok Kim argues that the fourth sprout, i.e. the sense of right and wrong cannot be understood as an emotion as it “denotes a markedly different psychological state than the other three sprouts do” (Kim 2014, 53). The objection is an interpretive one. However, it will likely be concerning to those who argue that Mencius’ account of moral judgments is an EAM. If the objection is successful, it would not only dethrone one of the four sprouts as an emotion, it would also provide for a non-emotion-based source of moral judgment. Allow me to elaborate on this point. If we can legitimately assume that for Mencius all moral judgments find their source in the four sprouts, Mencius can only be an EAM theorist (strictly so called) provided all four sprouts are emotions. If one of the four sprouts is not an emotion, then it opens up the possibility of moral judgments that are not based on emotions. As such, if Kim’s objection succeeds (i.e. the fourth sprout is non-emotive) and the fourth sprout does not in one way or another work in tandem or depend on the other three emotive sprouts, it may negatively affect Mencius’ credentials as an EAM theorist.

Kim essentially offers three reasons for his objection. How *shi* 是 and *fei* 非 is used in the *Mencius* seems to refer to (a) a *judgmental activity* rather than a cognitive feeling, (b) a *complex judgmental activity*, and (c) the *reason-backed attribution* of rightness and wrongness to particular situations (Kim 2014, 57). Kim relies on two passages from the *Mencius* to support (a) and (b). However, for brevity’s sake I will only consider one. King Hui of Liang asked Mencius why his country’s population is stagnating despite his benevolent policies. Mencius answers:

“‘Since your majesty likes war, let me explain by [an analogy to] war. Drums are thundering, weapons have already crossed, and [soldiers] are fleeing, having thrown away their armor, and dragging their weapons. One stops after [running] a hundred paces, and another stops after fifty paces. What would you think if the latter, having run fifty paces, laughed at the one who ran a hundred paces?’ [the king] said, ‘Unacceptable. He merely didn’t run a hundred paces, but this is also fleeing.’ [Mengzi] said, ‘If your majesty knows this, you shouldn’t wish that your subjects were more numerous than those of the neighboring countries... Your dogs and swine eat the foods [intended] for humans, but you don’t know to put constraints on this; there are people dying of famine on the roads, but you don’t know to open your granaries; and when people die, you merely say ‘It’s not me, [it’s a bad] year.’ How is this different from killing a person by stabbing him and then saying that it’s not me, [it’s] the weapon? Don’t blame the year, your majesty; then all the people under heaven will arrive at your place’” (Kim 2014, 54).

With the above and in support of (a), we see that King Hui rejects as unacceptable the action of the fleeing soldier who ran a hundred paces. Kim argues that:

“‘To make this point the king uses ... ‘*shi* 是 A *ye* 也 (This is A)’ – that is, ‘this is also fleeing’ (shi yi zou ye 是亦走也); and from the fact that he utters this sentence as a reason for his negative attitude toward the runaway soldier who laughed at another of his kind, we can see that this utterance of his – ‘This is also fleeing’ – is a judgment fully loaded with a commitment to truth rather than a mere feeling that the king may reject in the end as a mistaken thought” (Kim 2014, 54).

Kim’s argument here appears to be that for Mencius the sense of right and wrong cannot be emotive as it is a judgment that involves not only a perception of things but also an element of assent. One of his reasons for taking up this position is the way King Hui used the word *shi* 是 in response to Mencius’ question. There is at least one plausible reply to Kim and it is as follows: It appears plausible for one to reject his implication that emotions cannot involve judgments strictly so called, i.e. judgments that involve assent. As we have seen with Nussbaum’s TBA, emotions can involve judgments that are assented to. For Nussbaum (as discussed in chapter one), emotions involve judgments and they broadly have two stages, i.e. the appearance stage and assent stage. As such, the claim that King Hui’s judgment that is ‘fully loaded with a commitment to truth’ cannot by itself be a reason for one to think that this sense of right and wrong is non-emotive. If

so, we need not consider the sense of right and wrong as a non-emotion-based source of moral judgment.

It is worth noting that Kim does not argue that the element of assent is the only feature of the sense of right and wrong that counts against it being emotive. I say this to ward off the suggestion that Kim's argument that *zhi* is non-emotive is based solely on a flimsy reason that the psychological state involves assent. A plausible response to (a) alone would not suffice in adequately addressing Kim's objection. We would also have to adequately address (b) and (c). I now turn to (b), i.e. the sense of right and wrong is a *complex* judgmental activity and this counts against it as being emotive. In support of (b), Kim argues that:

“the phrase ‘*shi* 是 and *fei* 非’ seems to be used in [the above] passage *not only for making a simple judgment but also... for comparing more than two things and categorizing them according to their characteristic features...* Mengzi presses the king by asking him a rhetorical question: how is his blaming the year (i.e., the bad luck of the year) instead of blaming himself for his people's suffering any different from someone's blaming his weapon after killing a person using that weapon. In this question Mengzi uses *shi* 是 in the phrase ‘*shi he yi yu* 是何异于...’ (How is this different from...), and what Mengzi is doing here is challenging the king to provide a reason that could refute Mengzi's categorization of things... the word *shi* 是, which literally means ‘this’ but which I also take to carry in its wake ‘the characteristic features’ of ‘this thing’ is used by Mengzi in the categorization of or discrimination among things [emphasis added]” (Kim 2014, 53-54).

One can plausibly reply to Kim that he is being selective in the passages used to support his claim. Kim points out (in the above) that when Mencius uses the phrase ‘*shi he yi yu* 是何异于’, he means ‘how is this different from...’ He then argues that the word *shi* 是 does not just mean ‘this’ but also ‘the characteristic features of this thing’. Maybe *shi* 是 does carry such an implication in this specific context. I do not see a problem with this. But it is another step for him to argue that we can carry this implication into the context of Mencius' description of the fourth sprout, i.e. *shifei zhi xin* 是非之心 or the sense of right and wrong. No argument is put forward to support this claim and we cannot simply assume this claim. Kim argues that *shi* 是 means A in context X. But Kim failed to argue that *shi* 是's meaning A should be applied to context Y as well. There is an extra step required

to justify why *shi* 是's meaning in *shi he yi yu* 是何异于 should be applied to *shifei zhi xin* 是非之心. There are other parts of the text where *shi* 是 is used, Kim has to show why a particular usage of *shi* 是 is preferred over another. The following is an example from the *Mencius*:

“When Qi was afflicted with famine, Chen Zhen said, ‘The people of the state are all supposing that you, Master, will again request that the granary of Tang be opened for them, yet I am afraid that this may not be possible a second time.’

Mencius said, ‘To do it would be to do a Feng Fu [*shi wei Feng Fu ye* 是為馮婦也]. There was a man of Jin named Feng Fu who was good at seizing tigers. Later, he became a good scholar. But when he went out to the wilds, there was a crowd of people in pursuit of a tiger. The tiger took refuge in a mountain, where no one dared attack him, but when the people saw Feng Fu, they ran and welcomed him. Feng Fu bared his arms and descended from his carriage. The crowd was pleased with him, but the scholars laughed at him.”
(Mencius c 300 BC: 2009, 7B23)

We see here that *shi* 是 refers to ‘it’ and ‘it’ refers to the act of requesting to open the granary of Tang’. As we can see, the usage of *shi* 是 here does not imply ‘the characteristic features of this act of requesting’. Mencius compares the act of requesting with what Feng Fu experienced. He seems to be trying to convey that there are similarities between the act of requesting and Feng Fu’s experience, i.e. if I make the request the consequences would be similar to what Feng Fu had experienced. If this is correct, then *shi* 是 here implies not ‘the *characteristic features* of the act of requesting’ but ‘the *consequences* of the act of requesting’. The question here now is why should *shi* 是 in *shifei zhi xin* 是非之心 have a similar meaning to *shi* 是 in *shi he yi yu* 是何异于 but not *shi wei Feng Fu ye* 是為馮婦也. It appears then that for Kim’s argument to work he may have to not only show that *shi* 是's meaning A in context X should apply to context Y, he may also have to show that *shi* 是's meaning B in context Z does not apply to context Y. If so, until Kim provides us with such arguments, Mencian EAM theorists may not have to take his objection too seriously.

While my response to (b) challenges the manner Kim interprets the *Mencius*, there are other lines of responses. He argues that the fourth sprout is non-emotive as the psychological state involves the comparison and categorisation of things according to their characteristic features. One line of response is that even if we see the fourth sprout as a judgmental activity that involves comparison

and categorisation, this psychological state can still be one that involves emotion(s). And this would suffice in allowing the Mencian account of moral judgment to remain emotion-based. One way of going about this is to see the fourth sprout as a psychological state that works in tandem with the other three sprouts (as emotional states). The fourth sprout can draw on the 'insights' of the other three in the sense that the judgmental activity of this sprout can involve a comparison between differing emotions. The following passage from the *Mencius* can allow us to see how such a comparison might occur:

“Chunyu Kun said, ‘Is it a matter of ritual propriety that, in giving and receiving things, men and women should not touch one another?’

Mencius said, ‘This is according to ritual.’

‘If one’s sister-in-law is drowning, may one save her with his hand?’

‘If one’s sister-in-law were drowning and one did not save her, one would be a wolf. For men and women, in giving and receiving, not to touch one another is according to ritual. To save a sister-in-law from drowning by using one’s hand is a matter of expedience’

‘Now the whole world is drowning and yet you do not save it. Why is this?’

‘When the world is drowning, one saves it through the Way. If one’s sister-in-law is drowning, one saves her with one’s hand. Would you like me to save the world with my hand?’” (Mencius c 300 BC: 2009, 4A17).

With this excerpt, we can say that Mencius compared not touching one’s sister-in-law with saving one’s sister-in-law. On one hand, not touching one’s sister-in-law is in accordance with propriety. And deferential feelings (third sprout) can be what motivates this aversion to touch one’s sister-in-law. On the other, saving one’s sister-in-law can be motivated by one’s compassion (first sprout). One can then via the fourth sprout weigh one’s emotions (the first and third sprouts) in both situations to determine which takes precedence. In this case, we can plausibly say that Mencius determines that compassion trumps deference. This is why it is expedient to violate what is ‘proper’ to save one’s sister-in-law. It is in this sense that the fourth sprout works in tandem with the other sprouts (emotional states). The fourth sprout can still be an emotion-involving process, thereby allowing the Mencian account of moral judgment to remain as emotion-based.

With (b) addressed let us move on to (c), i.e. the reason-backed attribution of rightness and wrongness to particular situations is a reason that counts in favour of thinking of the fourth sprout as non-emotive. Kim relies on another passage from the *Mencius* to support (c):

“Chen Zhen 陈臻 asked: ‘Earlier [when you were] in Qi 齊, the king sent [you] a hundred *yi* 镒 of metal, but you didn’t accept it; but in Song 宋, [the king] sent seventy *yi* [of metal] and you accepted it, and while in Xue 薛, they sent fifty *yi* and you [also] accepted it. [However,] if your refusal on an earlier day were right, your accepting [a gift] recently should be wrong; and if your accepting it recently were right [*shi* 是], your declining it on an earlier day should be wrong [*fei* 非]. I’m afraid the Master should have acted consistently in this matter’. Mengzi said, ‘[My actions were] right [*shi* 是] in all of these cases. While staying in Song, I was about to go on a long journey; travelers are supposed to be [greeted] with a parting gift; and the message [accompanying the gift] was, ‘Presented as a parting gift.’ For what reason should I have refused it? While staying in Xue, I was apprehensive [of my safety], and the message was ‘I heard about your precautions [for your safety], so I sent this for [helping you purchase some] arms.’ For what reason should I have refused it? But as for [my staying] in Qi, I had no justification [for accepting a gift], and sending something [to a person] for no reason is bribing him. How could there be a nobleman who could be bought with a bribe?’” (Kim 2014, 56).

With the above, we see (as Kim argues) that the word *shi* 是 is understood as ‘right’ and it implies the taking up of a moral position that *x* is right. This brings us closer to how *shi* 是 is understood in *shifei zhi xin* 是非之心. And it can perhaps be conceded that the meaning of *shi* 是 in both contexts are the same or very similar. *Shi* 是 in both contexts involve the taking up of a moral position. I do not see a problem with this. Kim also takes the extra step of arguing that the use of *shi* 是 in the above passage implies the taking up of a moral position *that is expected to be substantiated with reasons*. I also do not see a problem with this. I draw the reader’s attention to Haidt’s social intuitionist model (Haidt 2001). He argues that “moral judgments appear in consciousness automatically and effortlessly as the result of moral intuitions” (Haidt 2001, 818). He then argues that “moral reasoning is produced and sent forth verbally to justify one’s already-made moral

judgment to others. Such reasoning can sometimes affect other people... [the hypothesis here is that] [b]ecause moral positions always have an affective component to them... reasoned persuasion works not by providing logically compelling arguments but by triggering new affectively valenced intuitions in the listener" (Haidt 2001, 819). In summary, Haidt's argument is that moral judgments are mainly affect-laden, but this does not prevent him from claiming that reasons can be proffered to convince others to take up one's moral position(s). Applying this point to Kim's argument, I can safely say that the reason-backed attribution of rightness and wrongness to particular situations is not necessarily a reason that counts in favour of thinking of the fourth sprout as non-emotive.

Furthermore, even if the fourth sprout does not work in tandem with the others, Kim's attempt to extend this comparison and categorisation implication in the above passage to how *shi* 是 is understood in *shifei zhi xin* 是非之心 is also problematic. Kim gives us no argument on why this is the case. Again, there is no presumption that such an implication should apply to *shifei zhi xin* 是非之心. In addition, how *shi* 是 is understood in *shifei zhi xin* 是非之心 and in the passage above is consistent with the view that all humans (barring any defect) have intuitions of right and wrong (their cognitive components of which are moral judgments) that can develop into more cognitively complex moral judgments. In other words, Kim's use of the above passage appears to pose no threat to a more traditional interpretation of the *Mencius* where the fourth sprout is understood as emotive.

Now I hope that with the above the reader can see that Kim's objection can be adequately addressed and the threat of a non-emotive source of moral judgment can be kept at bay. I argue further that even if it turns out that for *Mencius* *shilfei* is not an emotion, this would only mean that only one out of four sprouts is non-emotive. Moral judgments can still be grounded in the other three emotive sprouts. Given that there appears to be no need for an EAM theorist to claim that *every* moral judgment involves emotion, I take no issue in classifying *Mencius*' account as an EAM. Furthermore, even if for *Mencius* *shilfei* is not an emotion, it may not pose a threat to the aim of this chapter, i.e. to provide a *Mencian* account of moral judgment. A neo-Mencian (as opposed to *Mencius*) is still open to reject this claim as wrong or unnecessary, excise it from a proposed theoretical framework and salvage the good parts. For example, a neo-Mencian may

argue that all moral judgments are in one way or another grounded in the *three* remaining emotive sprouts. The fourth sprout can be regarded by the neo-Mencian as dross. Specifically, the neo-Mencian can reject *shilfei* as a sprout in the sense that it is not constitutive of Mencian baseline judgments. However, this does not mean that *shilfei* is completely excised from the neo-Mencian framework. *Shilfei*, if understood as a state that involves the comparison and categorisation of things according to their characteristic features, can still be involved in non-baseline judgments.

The Nexus Between Emotion and Non-baseline Judgment

With the nexus between baseline judgment and emotion set out and an objection against this nexus addressed, we can now consider the nexus between *non-baseline* judgment and emotion. We can understand this nexus if one considers carefully 1A7 where we see King Xuan asking Mencius “[c]ould someone like me protect the people?” (Mencius c 300 BC: 2009, 1A7). As discussed earlier in this chapter, Mencius answers this question by reminding King Xuan of his compassion toward the ox about to be slaughtered. After this reminder, Mencius argues that “[w]ith such a mind one has what it takes to be a true king” (Mencius c 300 BC: 2009, 1A7). As Wong argues:

Mencius “goes onto suggest that surely the king could extend the compassion he felt toward the ox to his people. To say that he could spare the ox from compassion but not spare his people, says Mencius, is like saying he can see the tip of an autumn hair but not a cartload of firewood... Mencius is trying to get him to *si* [reflect] on his motivations in trying to persuade him that he has the motivational wherewithal to be a true king. This part of the 1A7 passage contains a characteristic form of Mencian reasoning: reflecting takes the form of turning over in one’s mind certain elements of one’s past and present experience so that certain patterns or analogical associations emerge.” (Wong 2017, 235).

We can describe the analogical reasoning involved in 1A7 in the following manner: The king reacted with compassion in seeing the ox’s suffering and innocence. The king was motivated (via compassion) to spare the ox. The ox and the king’s subjects share similar characteristics. Just as the ox was suffering and undeserving of suffering, the king’s subjects are also suffering and undeserving of suffering. And in virtue of these characteristics shared between the ox and people, just as the king felt compassion and was motivated to spare the ox from suffering, he can also feel

compassion and be motivated to spare his people from suffering. The king's care for the ox can by analogy be extended to care for his people.

Apart from the above, Wong argues that in 1A7 "Mengzi urges [King Xuan] to compare the plight of the ox to that of his people and to compare his compassionate reaction to the ox to what ought to be the compassion he feels for his people... [this] reflection on the two cases involves the perception of relevant similarity and the recognition that one's reaction to [the ox]... brings, or should bring a relevant similar reaction to [the people]" (Wong 2002, 197). Here we see how the Mencian agent can 'extend' the repertoire of her moral judgments. In *Case A*, the Mencian agent experiences compassion toward *x*. The relevant features in *Case A* include the suffering and innocence of *x*. When the Mencian agent faces a novel situation (*Case B*), she can reflect on whether *Case B* has features relevantly similar to *Case A*. If *Case B* comprises features like the suffering and innocence of *y*, she can see that *Case B* has relevant similarities with *Case A*. And since *Case A* has the additional feature of *the suffering and innocence of x is a reason to help x*, the Mencian agent can also see that this feature should be 'extended' to *Case B*. This extension takes the form of a judgment in *Case B* that the suffering and innocence of *y* is a reason to help *y*. As McRae puts it succinctly for Wong, "the Mengzian agent judges that the case of the people is relevantly similar to the case of the ox, and so deserves the same moral response (compassion)" (McRae 2011, 593).

In sum, we can understand a non-baseline judgment of a Mencian agent in at least two ways. First, we can understand a non-baseline judgment as an *activity*, e.g. the Mencian agent's thinking of whether there is a reason to help *y* in *Case B*. This is the deliberative phase where analogical reasoning is involved. Second, we can understand a non-baseline judgment as a *state*. This state is the result of the Mencian agent's thinking of whether there is a reason to help *y* in *Case B*. If *Case A* is analogous to *Case B*, the non-baseline judgment here would be the state of seeing that there is a reason to help *y*. This is the executive phase where analogical reasoning is concluded, and a judgment (as a state) ensues. As we can see, a non-baseline judgment is emotion-based in at least two ways as well. The first way is that a non-baseline judgment as an *activity* relies on a baseline judgment (the cognitive component of emotion) as the starting point for analogical reasoning. The activity necessarily involves the Mencian agent making reference to a baseline judgment previously made. This being said, as an emotion's cognitive component is analytically and

phenomenologically distinct from its affective component, there appears to be no need for the Mencian agent (while engaging in analogical reasoning) to be in the relevant emotional state. The second way in which a non-baseline judgment is emotion-based is that such a judgment as a *state* is the product of an analogical reasoning process that necessarily refers to a baseline judgment previously made. As non-baseline judgments are ‘derived’ from baseline judgments (cognitive components of emotion), we see how non-baseline judgments have an emotive source.

Mencius’ Commitment to Non-baseline Judgments

With the nexus between emotion and non-baseline judgment set out, allow me to consider two objections. Let us start with the first objection. McRae objects to Wong’s interpretation of 1A7. She argues that:

“Mengzi is not encouraging the king to extend his moral response ‘from one case of right to another,’ since the king’s compassion for the ox is not, for Mengzi the correct response. Mengzi says ‘A gentleman is sparing (ai) with living creatures but shows not benevolence (ren) towards them.’ [7A45] The king’s moral responses are out of balance; he has compassion for an animal and indifference for his people. He has, in a sense, overshot the mark, since he gives a more intimate response than the occasion warrants” (McRae 2011, 593).

McRae’s objection, if successful, suggests that a Mencian account of non-baseline judgment needs to be developed from a passage other than 1A7 of the *Mencius*. Or non-baseline judgments can be formed in ways other than Wong’s interpretation of 1A7. This at the very least undermines Mencian a la Wong’s account of non-baseline judgments and, more importantly, puts into doubt whether Mencius is committed to non-baseline judgments in the first place. With this in mind, let us consider Wong’s response to McRae’s objection. He argues that McRae is wrong on her view that for Mencius compassion toward the ox is inappropriate: “[the king] could not bear the ox’s suffering having seen it, but (as Mencius himself believes), the ritual takes precedence over the ox’s suffering. And that is the reason why it was right to substitute the sheep. But once having accidentally seen the suffering of the ox, the king was right to spare it. This is what an exemplary person would do, says Mencius. And that is why the exemplary person stays away from the kitchen” (Wong 2015, 42). Now if Wong’s interpretation (as opposed to McRae’s) is correct, then

Wong's argument that 1A7 reveals a movement from one case of right to another via analogical reasoning still stands. Though this is one way of addressing the debate between McRae and Wong, my intention here is not to declare Wong as victor against McRae. Rather, I intend to use this debate as a segue into showing that there may be another plausible way of reading 1A7. And this reading can allow me to hold on to the position that 1A7 reveals how one can 'extend' the repertoire of moral judgments from baseline to non-baseline without settling the debate between Wong and McRae.

Wong's and McRae's disagreement lies essentially on whether for Mencius the king's sparing of the suffering ox was the correct response. The apparent assumption held by both of them is that Mencius in 1A7 is either explicitly or implicitly making a determination on the correctness of the response. But what if we eschew this assumption in our interpretation of 1A7? Allow me to explain. Mencius says to the king about his sparing of the ox from ritual slaughter and substituting it with a sheep that "[t]here is no harm in this. This was after all the working of humaneness- a matter of having seen the ox but not the sheep. This is the way of the noble person in regard to animals: if he sees them alive, then he cannot bear to see them die, and if he hears their cries, then he cannot bear to eat their flesh. And so the noble person stays far away from the kitchen" (Mencius c 300 BC: 2009, 1A7). Perhaps the message Mencius was trying to convey here was a descriptive (as opposed to a normative) one, i.e. (a) the king experienced compassion for the ox, (b) the king's compassion moved him to spare the ox, and (c) if the sheep was led to ritual slaughter and the king witnessed it he would also experience compassion toward the sheep and be moved to spare the sheep. There appears to be no need to read into 1A7 that Mencius thought the sparing of the ox was right or wrong. If I adopt this interpretation of 1A7, i.e. If Mencius in 1A7 was merely *describing* the king's experience of compassion and implicitly his perception of the ox's suffering as a reason for alleviating its suffering, I can continue to hold the position that 1A7 reveals to an extent how the Mencian agent can 'extend' the repertoire of moral judgments via analogical reasoning while avoiding the debate between Wong and McRae. In other words, we can dodge McRae's objection as it no longer applies to this proposed interpretation of 1A7.

One may object to the manner in which I dodge McRae's objection by saying that the *Mencius* is primarily didactic. It aims to teach the reader or in this case King Xuan something about how

he/she should feel, think and/or act. In this sense, we can understand the *Mencius* as a normative project (as opposed to a descriptive one). In reply, I accept that the *Mencius* is typically normative. However, it does not follow from this that we cannot understand parts of the *Mencius* descriptively. In the instance of 1A7, the normative (or didactic point) can be as follows: how one ‘extends’ the repertoire of moral judgments (from the seeing of the ox’s suffering as a reason to alleviate it to the seeing of one’s suffering subjects as a reason to alleviate their suffering) is an example of how the Mencian agent *should* ‘extend’ the repertoire of moral judgments. Here we see that the descriptive point (an example of how one ‘extends’ the repertoire of judgments) serves the normative point (the Mencian agent should ‘extend’ her repertoire of judgments in this manner). What is normative here is the manner in which one ‘extends’ her judgments and not the content of those judgments. 1A7 remains, by and large, normative. If so, this explanation is likely to assuage the concerns of those who regard the *Mencius* as typically normative.

As for the second objection, it is aimed at how the Mencian agent can arrive at non-baseline judgments. If a non-baseline judgment as a state results from an encounter with a novel situation and a process of treating like cases alike, the underlying assumption appears to be that the non-baseline judgment arrived at is a ‘new’ judgment (one that the Mencian agent has not made prior to the process of treating like cases alike). This assumption is apparent if we consider the process of treating like cases alike solely as a means to *increase* the repertoire of moral judgments. Allow me to explain. The process of treating like cases alike (roughly) goes like this: *x* has relevant characteristics *a*, *b*, and *c*. *x* also has characteristic *d*. *y* also has relevant characteristics *a*, *b*, and *c*. Barring any relevant dissimilarities between *x* and *y*, it is probable that *y* also has characteristic *d*. If we understand this process solely as a means to *increase* the Mencian agent’s repertoire of moral judgments, then it appears that the judgment *y* has characteristic *d* is necessarily a ‘new’ judgment.

This assumption, however, does not appear to be applicable in 1A7. Mencius’ account of analogical reasoning in 1A7 was prompted by the king’s question: “Could someone like me protect the people?” (Mencius c 300 BC: 2009, 1A7) and not ‘*should* I protect my people’. It appears from this that the king already takes for granted that he should protect the people. Even Wong concedes to this by arguing that the king “probably already knew that he had a moral reason to spare his people” (Wong 2015, 42). In 1A7, the king’s judgment that he should protect the people does not

appear to be a 'new' one. If so and as it appears, we cannot use 1A7 to support the claim that Mencius' position is that there are non-baseline judgments. Wong does not appear to be concerned by this as he is clearly aware that in 1A7 analogical reasoning may not have been used by Mencius to try "to teach the king anything new..." (Wong 2015, 42). Instead, analogical reasoning may have been used "to get the king to feel the reason to spare his people" (Wong 2015, 42).⁴⁴ But this *should* be a concern as the textual evidence in support of the cluster of claims on non-baseline judgments and how they are derived from baseline judgments appear to be lacking. The evidence put forward by Wong so far (3A5, 2A6 and 1A7) do not refer to reasoned judgments as described above.

There appears to be insufficient explicit textual support that for Mencius there are non-baseline judgments and such judgments are derived from baseline judgments via analogical reasoning. Instead of arguing that it is plausibly Mencius' stance that there are non-baseline judgments as described, I argue that we can understand this as a *neo-Mencian* stance. Non-baseline judgments can be neo-Mencian in that (a) it need not be a stance purported to be Mencius' despite being inspired by the *Mencius*, and (b) it is a stance put forward to fill a gap, i.e. to present a plausible emotion-based account of moral judgment. By making this move, there appears to be no longer a need to show that Mencius himself argued for an account of non-baseline judgments. If so, the evidentiary concern described above does not apply to a *neo-Mencian* stance. The neo-Mencian position here is that 1A7 reveals Mencius' *explication* of analogical reasoning used in the making of non-baseline judgments but not his *application* of analogical reasoning in the making of such judgments. Though the analogical reasoning in 1A7 may not be used by the king to increase his repertoire of moral judgments, it is this form of reasoning that can be used to do so.

There are other parts of the *Mencius* that reveal Mencius' explication of analogical reasoning, this includes 4A17 (see passage quoted earlier in this chapter). In this passage we see another instance of analogical reasoning at work. But instead of Mencius arguing that *Case A* has a relevant similarity with *Case B*, we see him arguing that *Case A* has a relevant *dissimilarity* with *Case B*.

⁴⁴ He further argues that in 1A7 "Mencius is trying to get the king to consciously reinstate his emotion of compassion through reflection on the analogies from the suffering of the innocent man to the suffering of the ox to the suffering of the people. He is reflecting on and reliving what he felt for the innocent man and the ox, and Mencius is hoping to get that feeling, the bodily readiness to respond and the motivational inclination, to flow to the king's people" (Wong 2015, 42).

Mencius applies analogical reasoning by arguing that saving one's sister-in-law is disanalogous from saving the world. There is a relevant dissimilarity between saving one's sister and the world. One can save one's sister-in-law with one's hand, but one cannot save the world with one's hand. As such, Chunyu Kun cannot use the case of suspending ritual propriety in saving one's sister-in-law to suspend ritual propriety to save the world. Other Mencian scholars also adopt this view that Mencius applies analogical reasoning in 4A17. Wong argues, "Mengzi grants that to save the life of one's drowning sister-in-law, one of course suspends the customary rule of propriety prohibiting the touching of one's sister-in-law. A rival philosopher, Chunyu Kun, wants to apply this idea of suspending the usual rules of propriety to save something else – the world! Mengzi replies that one saves one's sister-in-law with one's hand but cannot save the world with one's hand" (Wong 2002, 197-198).

In sum, there is sufficient material on analogical reasoning in the *Mencius* to support a *Mencian* account (as opposed to *Mencius'* account) of non-baseline judgments. Though there may not be sufficient textual evidence in support of how analogical reasoning is used in the making of non-baseline judgments, there is no need for such evidence to support our Mencian account. This is due to the aim of such a Mencian account, i.e. the account is put forward to fill a gap by presenting a plausible Eastern emotion-based account of moral judgment.

Broadening the Epistemic Base

With a Mencian account of moral judgment set out and some of its objections addressed, I am now able to compare this account with some Western EAMs discussed in the previous chapter. This comparative exercise is aimed at allowing the reader to identify the key distinguishing features of the Mencian account. I hope that by doing so the reader can see with more detail as to how the epistemic base for EAMs is broadened with the introduction of the Mencian account.

As a start, Mencian baseline judgments are similar to Zagzebski's level 1 judgments, i.e. the emotional state of seeing x as having a thick affective property (plus assent). They are similar in the sense that both Mencian baseline judgments and Zagzebski's level 1 judgments in one way or another involve perception. But as it appears, Mencian baseline judgments involve a far richer

range of emotional perceptions. For Zagzebski, being in the emotional state of compassion involves seeing x as pitiable. The level 1 judgment here appears to be restricted to ' x is pitiable'. As opposed to Zagzebski's account, the Mencian account of compassion need not be limited to the state of seeing x as pitiable. The cognitive component of compassion can also involve seeing x as innocent, seeing x as undeserving of punishment, seeing x 's suffering as a reason for acting in helpful ways. Furthermore, and as it appears, the level 1 judgment involved in compassion is *not* a judgment of culpability. Whereas Mencian baseline judgments 'stemming' from compassion can be a judgment of culpability, i.e. perceiving that x is innocent/undeserving of suffering.⁴⁵

We can further illustrate this rich range of perceptions that the Mencian account offers if we compare the Mencian perception of x 's suffering as a reason for helping with Zagzebski's "judgment that I ought to help her" that "may be an expression of emotion" (Zagzebski 2004, 149). Zagzebski argues that (a) "[i]f I see her as suffering, I may see her as pitiful, and I may see a potential act in response as obligatory" (Zagzebski 2004, 149), (b) this "sense of obligation... arises between two emotions – the primary emotion of detecting the thick affective property of some situation [in our case compassion], and the emotion of guilt or shame when the agent does not respond to the situation by acting in a certain way" (Zagzebski 2004, 151), and (c) this "sense of the obligatory is the emotion one has when one considers a potential act (or omission) by oneself that would be an attack on the self" (Zagzebski 2004, 159). What I take from her argument is that the seeing of the alleviation x 's suffering as obligatory is an emotion by itself. And though this sense of obligation arises 'in between' compassion and guilt/shame, it is clearly distinct from them. As it appears, Zagzebski's account differs from the Mencian account. To the contrary, in the Mencian account, the seeing of x 's suffering as a reason for helping does not arise between compassion and guilt/shame. This perception is not a separate emotion by itself. Rather, it 'stems' solely from the cognitive aspect of compassion. More specifically, for the Mencian, to see x 's suffering as a reason to act in helping ways is paradigmatically constitutive of the emotional state of compassion.

⁴⁵ It is worth noting that we can speak of Mencian compassion as both a disposition and an emotional state. When Mencius argues that "[a]ll human beings have a mind that cannot bear to see the sufferings of other..." (Mencius c 300 BC: 2009, 2A6), he is referring to what he considers to be an inborn disposition to feel compassion. When this disposition actualises, we can then speak of Mencian compassion as an occurrent emotional state. In this paragraph, I am referring to Mencian compassion as an emotional state (as opposed to a disposition).

This difference between Zagzebski's and the Mencian account is both unimportant and important. Some may see the notion that Zagzebski accounts for a *separate* emotion where an agent can see the alleviation of *x*'s suffering as obligatory (when compared to the Mencian account) as *unimportant* in the sense that both accounts still proffer (albeit in different ways) an emotion-based judgment of obligation. However, from another perspective, others may see the difference as *important* in that the Mencian account presents a more direct account of emotion-based judgments of obligations. As the seeing of *x*'s suffering as a reason to help for the Mencian is by itself paradigmatically constitutive of compassion, the Mencian account does not require two prerequisite emotions (in this instance compassion and guilt/shame) for an agent to experience an emotional sense of obligation. The directness of the Mencian account can, at the very least, showcase an example of the range of options available to fence-sitters who might consider emotion-based accounts of moral judgments as plausible.

The Mencian account also offers a distinctive perspective on how emotion is constitutive of baseline/moral judgments. As argued in the preceding chapter, Prinz and Zagzebski (despite their differences) are committed to the claim that certain types of moral judgments are constituted by emotion. For Prinz, the compound psychological state of anger at pick-pocketing or guilt at incest (e.g.) is a kind of moral judgment. The emotion (anger or guilt) *as a whole* is constitutive of moral judgment. But this does not appear to be the case for the Mencian account. As argued above, baseline judgments are not constituted by an emotion as a whole but by the emotion's cognitive component alone. For Zagzebski, an emotion is a unitary state in that its cognitive and affective aspects cannot be phenomenologically and analytically pried apart. And such an emotional state (the state of seeing *x* as having a thick affective property) plus assent can also be a moral judgment (level 1 judgment). If so, it is the emotional state *as a whole* that is constitutive of the level 1 judgment.⁴⁶ This again is a point of departure for the Mencian account. The Mencian account puts forward the claim that compassion, shame, deference and *shifefi* have cognitive and affective

⁴⁶ It is worth noting that, while Zagzebski allows for level 2 judgments (seeing *that x* has a thick affective property), level 2 judgments are not emotional states. The point I wish to make here is that while for Zagzebski the cognitive and affective aspects *can* be pried apart in level 2 judgments, the cognitive and affective aspects *cannot* be pried apart in level 1 judgments.

components. These can be phenomenologically and analytically pried apart. And it is only the cognitive component of emotion that constitutes baseline judgments.

In addition, Mencian baseline judgments differ from Zagzebski's level 2 judgment, i.e. the judgment state of seeing *that x* has a thick affective property plus assent. Level 2 judgments from a phenomenological perspective involve the thought that *x* is pitiable (e.g.) *without pity being felt*. Whereas baseline judgments (as cognitive components of emotions) *are typically accompanied by the feelings and motivational tendencies of these emotions*. Furthermore, baseline judgment is not a judgment without emotion (as it comprises the cognitive aspect of the emotion), nor does it involve imagining being in an emotional state. Whereas Zagzebski argues that "[w]hen I make a judgment without emotion [a level 2 judgment]... it is likely that I must still be able to imagine seeing something as lovable... but even imagining requires a faint copy of the feeling of love... in my mind" (Zagzebski 2004, 168). Further to this point, level 2 judgments are more similar to Mencian non-baseline judgements when compared to baseline judgments. Non-baseline judgments can be judgment states resulting from an analogical reasoning process. As such, they can be 'judgments without emotion'. The Mencian agent can as a result of analogical reasoning judge that she should have feelings of compassion toward *y* but need not actually have such feelings. But despite this similarity with level 2 judgments, non-baseline judgments can still be as emotionally charged as baseline judgments. I will discuss this with more detail in the final paragraph of this section. But it would suffice to note now that this feature is a point of departure from Zagzebski's level 2 judgments. Level 2 judgments are not emotional states like level 1 judgments. Hence, level 2 judgments cannot be emotionally charged.

The pluralism of the Mencian account has also some similarities with and differences from that of Prinz. I will consider some in the rest of this section. As argued in the previous chapter, there are three types of moral judgments for Prinz. The first type of moral judgment for Prinz can be an utterance, i.e. the voicing of one's unactivated/activated moral sentiment. But unlike Prinz's account, such or any other sort of account of verbalized judgments is absent from the Mencian account. More specifically, while the *Mencius* records verbalized sayings/judgments of Mencius, there is no *explicit* account in the *Mencius* on the nexus between the emotive sprouts and verbalized judgments. This being said, the neo-Mencian can say that the utterance that '*x* is right' can be a

verbalized expression of a baseline or non-baseline judgments (both of which rest on emotional experience). Allow me to consider an example. In 4A17 above, we saw Mencius say, “If one’s sister-in-law were drowning and one did not save her, one would be a wolf”. The neo-Mencian can reframe this utterance in several ways, including ‘saving one’s drowning sister-in-law is right’. The neo-Mencian can now also say that this utterance can be a verbalized expression of the emotional state of compassion. More specifically, the utterance can be the verbalized expression of the seeing of one’s drowning sister-in-law as a reason to act in helping ways. This position is plausible if we consider carefully Mencius’ claim that one who does not save her sister-in-law is like a wolf. The underlying assumption of Mencius here appears to be that wolves act in inhumane and cruel ways. While this assumption is probably wrong, the point here is that Mencius could have instead said, ‘if one’s sister-in-law were drowning and one did not save her, one would lack *compassion*’. What Mencius (wrongly) finds inhumane and cruel in wolves is juxtaposed against what he finds in all humans, i.e. the sprout of compassion. There is as such a similarity between Prinz’s account and the Mencian’s, i.e. both have accounts of moral judgments as utterances. And these utterances in one way or another rest on emotional experience. For Prinz, such utterances verbalise either activated/unactivated moral sentiments. Whereas for the Mencian account such utterances verbalise either baseline/non-baseline judgments.

The second type of moral judgement for Prinz is (e.g.) the psychological state of reacting with guilt/anger toward *x* (activated moral sentiment). Now such judgments are similar to Mencian baseline judgments at least in one respect. Similar to baseline judgments, activated moral sentiments can be spontaneous and occur without deliberation. For the Mencian account, when I see a child suffering, I can without further deliberation see via compassion that alleviating the child’s suffering is the right thing to do. Similarly for Prinz, when I see someone stealing, I can without deliberation react with anger toward the act of stealing. But though baseline judgments and Prinz’s activated moral sentiments have the shared characteristic of spontaneity, there is one key difference I would like to point out. A moral sentiment for Prinz is understood as “a standing belief” in one’s long-term memory and anger toward stealing (e.g.) as a “particular emotional manifestation of that sentiment” (Prinz 2007, 96). But a baseline judgment need not involve any such sentiment/belief. I can see/judge that alleviating the child’s suffering is the right thing to do without having such a belief. As Wong argues, “the cognitive component [of the emotion] need not deliver judgments or beliefs *in any full-blooded sense*. A child can feel alarm and distress over a

favorite doll's broken arm. She may perceive the doll to be suffering; but does she believe it? Such phenomena indicate that the cognitive component can sometimes remain at the level of 'seeing something x as if it were y ' that precedes and often, but not always, turns into belief' [emphasis added]" (Wong 2015, 29). This point is important for the Mencian account as it leaves room for instances where one makes a baseline judgment but does not assent to it. I can (e.g.) judge via compassion that the alleviation of my child's suffering while visiting the dentist is the right thing to do. But it does not necessarily mean that I believe that alleviating her suffering in this instance would be the right thing to do.

The third type of moral judgment for Prinz is a psychological state of being disposed to react with guilt/anger toward x (unactivated moral sentiment). Such an unactivated moral sentiment is similar to a Mencian non-baseline judgment in at least one respect. Both judgments can be 'cool' and unemotional. For Prinz, I can think that x is wrong without reacting with anger/guilt toward x . And the thought that x is wrong is genuine as long as I have a standing belief in my long-term memory that x is wrong, and I am disposed to react with anger/guilt toward x . Similarly, for the Mencian account, a non-baseline judgment as a judgment state resulting from an analogical reasoning process can be 'cool' and unemotional. Though this reasoning process relies on baseline judgments as its starting point and necessarily involves the Mencian agent making reference to such a judgment, there appears to be no need for the Mencian agent to be in any emotional state when a non-baseline judgment is made. Despite this similarity, there is also one key difference. For Prinz, the genuine thought that x is wrong (as an unactivated moral sentiment) is necessarily unemotional. Whereas, a non-baseline judgment can be unemotional but need not be. I may have a non-baseline judgment that I should extend my compassionate response from *Case A* to *Case B*. This judgment can be unemotional in the sense that I need not have a compassionate response toward *Case B* for me to judge that I *should* extend my compassion to *Case B*. This being said, I may also have an *emotionally-charged* non-baseline judgment where as a result of analogical reasoning I can see via compassion that in *Case B* the suffering and innocence of y is a reason to help y and be motivated to act accordingly. In other words, analogical reasoning can not only lead the Mencian agent to judge that she should have a certain emotional response toward *Case B*, it can also result in her actually having this emotional response toward *Case B*. This Mencian stance is similar to Wong's argument that Mencian extension "involves both a growth in *judgment and feeling* [emphasis added]" (Wong 2002, 196-197). There is however one key difference between Wong's

position and the Mencian account put forward. For Wong, he puts this argument forward as an interpretation of the *Mencius* (in particular 1A7). As for me, I remain neutral on whether Mencian extension as described is a position actually held by Mencius.

The Water and Sprout Metaphors

With some key differences between the Mencian account and some Western EAMs outlined, we can ask at this point how does the Mencian account of moral judgment help explain moral unanimity and diversity. To answer this question, I argue that the Mencian account has two readily accessible metaphors useful in explaining moral unanimity and diversity. They are the water and sprout metaphors. These metaphors reveal the claim that we have inherent tendencies to judge in certain ways. This as we shall see can help in explaining some aspects of moral unanimity. The metaphors also reveal the claim that our inherent tendencies are not impervious to culture, environment and personal reflection. This in turn can help in explaining some aspects of moral diversity.

Without further ado, let us begin with the water metaphor. In 6A2, we see a debate between Gaozi and Mencius. The former argues that “[h]uman nature is like swirling water. Open a passage for it in the east, and it will flow east; open a passage for it in the west, and it will flow west”. Mencius rejects this and argues instead that “... human nature is like the downward course of water. There is no human being lacking in the tendency to do good, just as there is no water lacking in the tendency to flow downward. Now, by striking water and splashing it, you may cause it to go over your head, and by damming and channelling it, you can force it to flow uphill. But is this the nature of water? It is force that makes this happen” (Mencius c 300 BC: 2009, 6A2). Mencius’ response to Gaozi that water naturally flows downward can be known as the water metaphor. Wong argues that this metaphor “imparts a dynamic quality to human nature by characterizing it in terms of direction of movement, of change; second, this direction is toward the ethical... [and] [j]ust as water can be forced upward by splashing it or damming it up, so human nature can be channelled away from goodness” (Wong 2015, 162-163, similarly Slingerland 2003, 152 and Virag 2017, 113).

At this point, we can see how the water metaphor can help explain moral unanimity and diversity. First, 6A2 emphasizes that we have inherent tendencies. This point is obvious when we understand it as Mencius' response to Gaozi who argues that human nature is like swirling water with no set direction. To the contrary, human nature for Mencius is set to flow downwards. We have tendencies that inherently 'move downwards'. For example, we have a tendency to react with compassion if we see a child about to fall into a well. Via compassion, we also have a tendency to judge the child's future suffering as a reason to act in helping ways. This Mencian recognition of inherent human tendencies in 6A2 can assist us in explaining why there is some level of moral unanimity. If humans regardless of their cultural environment have such compassionate tendencies, this can explain (e.g.) why there is some level of cross-cultural consensus that we should help those who are suffering. Second, 6A2 reveals that such inherent tendencies are not impervious to change. Though water flows downwards, it can be dammed up and flow against the downward direction. Analogously, though we have inherent tendencies to act or judge in particular ways, these tendencies can be curbed by cultural practices. The inherent-ness involved here is comparable to what Gary Marcus argues to be innate: "Nature provides a first draft, which experience then revises... 'built-in' does not mean unmalleable; it means 'organized in advance of experience'" (Marcus 2004, 34, 40). The first draft is comparable to the set direction of certain tendencies. And the draft's openness to revision is comparable to the openness of these tendencies to be led in different directions. While the set direction of these tendencies can explain moral unanimity, their susceptibility to be led in different directions can explain moral diversity. For example, though we have a tendency to judge via compassion that suffering is a reason to act in helping ways, cultural practices can 'dam up' this tendency. They can carve out exceptions to this judgment. The caste system in India and the cultural practices associated with it can serve as an example. Whitley Kaufman, for example, argues that "[r]eportedly, the untouchables in India originally resisted Mother Theresa's [sic] attempts to improve their plight, as these might interfere with their karmic progress" (Kaufman 2007, 559). The practice here then is to leave the untouchables (despite their suffering) as they are. The rationale behind this practice may be that any attempt to alleviate their suffering would interfere with their karmic progress. Such a practice, as a result, may suppress or 'dam up' any initial judgment via compassion that an untouchable's suffering is a reason to act in helping ways.

With the water metaphor and its usefulness in explaining moral unanimity and diversity outlined, I now do the same with the sprout metaphor. We can find Mencius' description of the sprout metaphor in 2A6 discussed above. For the reader's convenience, I briefly repeat it here. For Mencius, there are four sprouts, i.e. compassion, shame, deference, and the sense of right and wrong. Mencius also argues that "[h]uman beings have these four sprouts just as they have four limbs" (Mencius c 300 BC: 2009, 2A6). Just as it is 'natural' for humans to have four limbs, it is also 'natural' for humans to have the four sprouts. For Mencius, humans as such have inherent tendencies to react with compassion, shame, deference, and a sense of right and wrong when particular circumstances arise. We can see that the sprout metaphor is comparable with the water metaphor in that both metaphors involve the claim that humans have certain inherent tendencies. Both the set flow of water and the growth of sprouts in certain ways signify the inherent tendencies that for Mencius humans possess. Other Mencian scholars also recognise this feature of the sprout metaphor. Edward Slingerland, for example, argues that "Mencius's primary metaphor is the dynamic 'sprout,' which has a *natural direction* and motive force of its own [emphasis added]" (Slingerland 2003, 131). But though the water and sprout metaphors share this common feature, as we have seen in 2A6, the latter is more detailed on what these tendencies are.

Like the water metaphor, the sprout metaphor allows for the view that cultural practices can influence the course of our inherent tendencies. Phillip Ivanhoe argues that the Confucian *rites* "guide and support their development until the sprouts are able to stand on their own... A healthy, vital specimen which grows undamaged will follow the course and assumes the shape described by these supports" (Ivanhoe 1990, 94). In short, the trellises of Confucian rites can guide the growth of our inherent tendencies. For example, we may have a tendency to see our parents as worthy of respect and see this as a reason to act in respectful ways. But we may not know exactly how to do so. We can seek recourse in the cultural practices that outline how we should treat our parents while they are alive and when they have passed on. Upon 'imbibing' the Confucian rites, we can not only see (via deferential feelings) our parents as worthy of respect but also see this as a reason to act in respectful ways *as set out by the rites*. We see this lack of certainty on how to be respectful and the Confucian solution that points to the rites clearly in the *Analects*:

"Meng Yizi asked about filial piety. The Master replied, 'Do not disobey.'

Later, Fan Chi was driving the Master's chariot. The Master said to him, 'Just now Meng Yizi asked me about filial piety, and I answer, 'Do not disobey.'"

Fan Chi said, 'What did you mean by that?'

The Master replied, 'When your parents are alive, serve them in accordance with the rites; when they pass away, bury them in accordance with the rites and sacrifice to them in accordance with the rites.'" (Confucius c 500 BC: 2003, 2.5)

Now the claim that Confucian rites can guide the growth of our inherent tendencies can be generalized. Instead of limiting the guidance of our tendencies to Confucian rites, one can say that cultural practices in general guide our tendencies. I may via the sprout of deference see Pope Francis as worthy of respect and see this as a reason to act in respectful ways. But should I bow before him? Or should I just shake his hand? Or should I genuflect? Existing cultural practices as 'trellises' can offer some guidance on how I should treat the pope with respect. I can pick up from the Catholic tradition on what to do when I meet the pope. I can as a result see the pope as worthy of respect and see this as a reason to address him as 'his holiness', genuflect before him and kiss his ring.

Apart from the influence of cultural practices, the sprout metaphor can also reveal how environmental factors influence our inherent tendencies. Wong argues that "[u]nlike the water metaphor, the sprout metaphor accords a crucial role for environmental factors. Sprouts need fertile soil and adequate sun and water" (Wong 2015, 163). And Flanagan argues that "[i]f [the sprouts] receive suboptimal nourishment, they will grow some; and if they are not nourished at all, they will not grow" (Flanagan 2014, 35). So instead of just staying out of the way of the set direction of water (or our inherent tendencies), we have to make an effort to nurture our sprouts (or inherent tendencies). We need to make sure that our sprouts have adequate sun and fertile soil. We need to keep them away from grazing sheep and cattle (Mencius c 300 BC: 2009, 6A8). 2A6 reveals that the sprouts or certain human tendencies are inherently directed toward virtue. And this for Mencius is the goal in nurturing the sprouts. This may be the case, but this is not my point of emphasis. The question here is not how does the sprout metaphor provide guidance in the attainment of virtue. Rather, the question is how does the sprout metaphor help explain moral unanimity and diversity. As such, my point of emphasis is that the sprout metaphor reveals to us

how environmental factors can influence our inherent tendencies. We see a glimpse of how environmental factors can have an impact in the growth of our sprouts in 6A8:

“Mencius said, ‘The trees on Ox Mountain were once beautiful. But being situated on the outskirts of a large state, the trees were cut down by axes. Could they remain beautiful? Given the air of the day and the night, and the moisture of the rain and the dew, they did not fail to put forth new buds and shoots, but then cattle and sheep came along to graze upon them. This accounts for the barren appearance of the mountain. Seeing this barrenness, people suppose that the mountain was never wooded. But how could this be the nature of the mountain?’” (Mencius c 300 BC: 2009, 6A8)

As it appears, our sprouts or inherent tendencies can be ‘cut down’ or ‘grazed by cattle’ which may not be within our control. One may have the tendency to judge via compassion human suffering is a reason for acting in helping ways. One may also have the tendency to judge via *shilfei* that eating human flesh is wrong. But these tendencies can be ‘cut down’ or ‘grazed upon’ by the harsh environmental conditions that one is in. Let us consider an example. The Aztecs were known for their human sacrifices and cannibalism. And Michael Harner argues that the cannibalism that took place were necessary for the “satisfaction of essential protein requirements” due to “conditions of environmental circumscription”, “high population pressure”, “an emphasis on maize agriculture”, a “Mesoamerican ecological area characterized by substantial wild-game degradation” and “the lack of a domesticated herbivore” (Harner 1977, 132). If Harner is correct and by applying the sprout metaphor to this case, we can say that the Aztecs may have had a ‘once beautiful Ox Mountain’ where they had tendencies to judge that human suffering is a reason for them to act in helping ways and that cannibalism is wrong. But such tendencies were subsequently cut down by the axemen of the environmental conditions described above. The Aztecs appear to have carved out exceptions to the injunction of helping those who are suffering and the prohibition on eating human flesh.

The sprout metaphor also reveals another way where our inherent tendencies can be ‘grown’, i.e. our inherent tendencies can be regulated in one way or another via personal reflection. We can already see a glimpse of this in the water metaphor when Mencius praises of Yu’s irrigation methods:

“Bai Kui said, ‘When it comes to water control, I am better than Yu.’

Mencius said, ‘Sir, you are mistaken. Yu’s control of water followed the Way of water. Therefore, he channelled it into the four seas, whereas you, sir, channel it into neighboring states. When water overflows its course, it is called a deluge, and a deluge is a flood – something that a humane man detests. Sir, you are mistaken.’” (Mencius c 300 BC: 2009, 6B11)

Slingerland’s interpretation of this is that for Mencius “[w]ater naturally flows to the sea; in regulating it, Yu simply helped it to its natural ‘home’...” (Slingerland 2003, 152). In other words, in 6B11 water has a set direction and we can help steer it in that direction if we would want to do something about it. We can understand Mencius as effectively saying that we *should* regulate water/our inherent tendencies toward the sea and not in the opposite direction. Given that the point of this section is to show how these metaphors can help in explaining moral unanimity and diversity, I would not assess this normative claim. What I would like to tease out of 6B11 is that we can regulate the flow of water/our inherent tendencies in one way or another. We can regulate it the Yu way or the Bai Kui way. We can understand Yu’s way as a way of ‘extending’ our tendencies to judge in certain ways. And Bai Kui’s way to ‘restrict’ our tendencies to judge in certain ways. We can for example upon personal reflection regulate the flow of our inherent tendency to judge that *a child’s* suffering is a reason to act in helping ways toward the judgment to “protect *all* within the four seas [emphasis added]” (Mencius c 300 BC: 2009, 1A7). We have already seen how this reflection might work when we move from baseline judgments to non-baseline judgments. 6B11 also reveals that we can steer water in another direction. Our inherent tendencies might be set toward a certain direction, but like water its course can be altered. While we might have the inherent tendency to judge that a child’s suffering is a reason to act in helping ways, we can upon reflection reject this judgment on grounds that the child is not my kin (e.g.).

The claim that we can regulate our inherent tendencies in one way or another via personal reflection can be seen even more vividly with the sprout metaphor:

“Always be doing something, but without fixation, with a mind inclined neither to forget nor to help things grow. One should not be like the man of Song. There was a man of Song, who, worried that his seedlings were not growing, pulled them up. Having done so, he

returned home wearily, telling people, 'I am tired today – I have been helping the seedlings to grow.' When his sons rushed out to have a look, they found all the seedlings withered. There are few in the world who do not try to help the seedlings grow. Those who believe there is no way to benefit them neglect the seedlings and do not weed them. Those bent on helping them to grow pull them up, which is not only of no benefit but, on the contrary, causes them injury" (Mencius c 300 BC: 2009, 2A2)

Here we see Mencius giving advice similar to Aristotle's golden mean. Yes, we should put effort into growing our sprouts. But we should neither put too little effort into it nor too much. These extremes would kill the sprouts. Again, I do not wish to assess Mencius' normative claim. Instead, there are two points that I wish to emphasize here. First, we can be involved in the nurturing of the sprouts. We can for example water them or pull them out. Second, we can reflect on how we would want to be involved in the nurturing of the sprouts. The man of Song wants the seedlings to grow, and he *thought* he could do so by pulling them out. Again, while we might have the inherent tendency to judge that my child's suffering is a reason to act in helping ways, we can (e.g.) upon reflection see that there is no relevant dissimilarity between my child and other children and 'water' this tendency by extending it to non-kin. And this watering process can take the form of the analogical reasoning process discussed above.

In sum, the water and sprout metaphors reveal that we have inherent tendencies to judge in certain ways and cultural practices, environmental factors and personal reflection can influence these tendencies. The point that we have inherent tendencies can help explain moral unanimity. And the point that such tendencies are not impervious to cultural, environmental and personal reflective influences can help explain moral diversity. Of course, there are still parts to be 'flesh out' especially on how do these influences interact and subsequently shape our inherent tendencies. For example, personal reflection can influence our inherent tendencies, but personal reflection is also likely to be influenced by cultural and environmental factors in the first place. In addition, though environmental factors may influence our inherent tendencies directly, it may also be that such factors first influence our cultural practices and our practices in turn influence our inherent tendencies. Or it may be that environment and culture simultaneously and independently influence our inherent tendencies. As the reader can see, there is still much to be explored. But

what is clear is that our exploration of a Mencian account of moral judgment thus far gives us unique access to metaphors helpful in explaining moral unanimity and diversity.

Interlude

With the above, we have seen how Mencian baseline and non-baseline judgments have an emotive source and how this Mencian account distinguishes itself from some Western EAMs. We saw as well how the Mencian account of moral judgment can adequately respond to Kim's objection that *shil/fei* is not an emotive sprout. This response is a novelty of this chapter, where a new argument has been produced to address the objection. We saw also how McRae's interpretation of 1A7 that put into doubt Mencius' commitment to non-baseline judgments can be addressed. I did so mainly by identifying and rejecting an assumption in McRae's interpretation. By adopting a neo-Mencian stance, I have also addressed the issue of there being insufficient explicit textual support for Mencius' commitment to non-baseline judgments. In addition, we have also seen a novel take on the water and sprout metaphors. Instead of using the water metaphor to say that allowing water to flow downwards (or in its inherent course) is 'good', and damming up water (or restricting water's inherent course) is 'bad', I use it to say that just as water inherently flows downward, humans have certain inherent emotional tendencies, and just as water's inherent flow can be 'dammed up' in various ways, our inherent emotional tendencies can be 'dammed up' in various ways. Instead of using the sprout metaphor to say that our emotive sprouts *should* be nourished in certain ways, and *should* be exposed to certain environmental factors, I use it to say that our emotive sprouts *can* be nourished in certain ways, and *can* be exposed to various environmental factors. As we can see, I have 'excised' the normative implications of the water and sprout metaphors. And this 'excision' places the metaphors in a unique position to explain aspects of moral unanimity and diversity. With this in mind, we can now consider how EAMs *in general* can explain moral unanimity and diversity and how their ability to do so is a brownie point in their favour.

Chapter Four

“A general law exists that has been made, or at least adopted, not only by the majority of such or such people, but by the majority of all men. This law is justice” (de Tocqueville 1835: 2010, 410)

“Justice, like finery, is dictated by fashion” (Pascal 1670: 1995, 25).

Explaining Moral Unanimity and Diversity

In chapter two, we reviewed some arguments in favour of EAMs. And in chapter three, we saw how an Eastern EAM is uniquely positioned to explain moral unanimity and diversity. In this chapter, I start by arguing that moral unanimity and diversity are salient features of morality. I will then argue for an EAM model that can explain such salient features. This is the novel contribution of this chapter. In the philosophical literature thus far, there has been no detailed discussion on the use of emotion-based accounts of moral judgments to explain (instances of) moral unanimity and diversity. Such a discussion, however, is warranted in at least one respect, i.e. if emotion-based accounts of moral judgments can explain the phenomena of moral unanimity and diversity, its explanatory power can serve as a reason in favour of the plausibility of emotion-based accounts of moral judgments. To address this gap, the EAM model I argue for in this chapter can plausibly explain moral unanimity by ‘clustering’ (instances of) pancultural moral conventions that span across time and space under certain pancultural emotions. The model can also explain moral diversity by illustrating how the malleability of (the objects of) certain pancultural emotions can result in the presence of moral conventions in one community and their absence in another.

Moral Unanimity

In this section, I aim to make clear two points. First, moral unanimity is a salient feature of morality in the sense that there are clear instances of shared moral principles and/or practices across

communities. Second, it is plausible that community *A* developed moral principles and/or practices shared by community *B* independently from each other.⁴⁷

Clear instances of moral unanimity. We can see many instances of moral unanimity. I will provide two examples of shared moral principles and/or practices. On the golden rule, we see myriad cultures (some of which may have been insulated from each other) making very similar moral pronouncements. In the *Book of Leviticus*, the author writes "... you shall love your neighbour as yourself" (Leviticus 19:18). In the *Analects*, Confucius says: "Do not impose upon others what you yourself do not desire" (Confucius c 500 BC: 2003, 15.24). In the *Mahābhārata*, the following advice is given: "... one should live with restrained soul, giving his attention to Virtue most. One should also behave towards all creatures as he should towards himself" (The Mahabharata c. 400 AD: 1890, CLXVII, 545-546). In the Pahlavi texts of Zoroastrianism, we see the following teaching: "... not to do unto others all that which is not well for one's self..." (Pahlavi Texts c. 250 AD: 1901, XIII: 29). In the *Sentences of Sextus*, we see the following remark: "What you do not want to happen to you, do not do it yourself either" (The Sentences of Sextus, 179). On alleviating suffering, it will not be too far-fetched to say that all communities now and before consider it morally obligatory/praiseworthy to help those in need. In the Book of Deuteronomy, the author writes: "When you reap your harvest in your field, and have forgotten a sheaf in the field, you shall not go back to get it; it shall be for the sojourner, the fatherless, and the widow..." (Deuteronomy 24:19). This pronouncement resembles the action of Australian aborigines: "In the Dalebura tribe a woman, a cripple from birth, was carried about by the tribes-people in turn until her death at the age of sixty-six." ... "They never desert the sick" (C. S. Lewis 1944: 2001, 97). We see similar actions in Native Americans as well: "You will see them take care of... widows, orphans, and old men, never reproaching them" (C. S. Lewis 1944: 2001, 97).

Independent development of moral principles/practices. As the reader would notice, I drew the above excerpts from different geographical locations and they belong to a time where the effects of globalism are minimal relative to the world we now live in. I sourced the texts and practices from

⁴⁷ This second point is important as it would accentuate how EAMs are uniquely positioned to explain moral unanimity. This would be made clearer in the next section. At this point, it would suffice to note that EAMs (with an evolutionary story) can explain how shared moral principles and/or practices across cultures could have independently developed.

ancient Israel, China, India, Persia, Greece, Australia, and North America. It is plausible that these communities were largely insulated from each other in the sense that contact between each of these cultures would have been minimal. And as for the Australian and Native American cases, there was no contact with the other cultures mentioned. As such, it would seem unlikely that principles of and/or practices involving the golden rule and alleviating suffering were spread from one community to another in the way that Catholicism, for example, was brought to South America by the Conquistadors. It seems plausible that these communities developed the principles and/or practices independently from each other. Though separated in time and space, they coincide in their approval of these principles and/or practices.

It should be made clear that the above is not a denial of the fact that ancient communities may have been influenced by each other. As Joseph Campbell argues:

“Persian mythology is rooted in the common Indo-European system that was carried out of the Aral-Caspian steppes into India and Iran, as well as into Europe. The principal divinities of the earliest sacred writings (Avesta) of the Persians correspond very closely to those of the earliest Indian texts [Vedas] ... Early in the first millennium B.C., Persian belief was reorganized by the prophet Zarathustra (Zoroaster) according to a strict dualism of good and evil principles, light and dark, angels and devils. This crisis profoundly affected not only the Persian, but also the subject Hebrew beliefs, and thereby (centuries later) Christianity...” (Campbell 1949: 2004, 322).

As we can see, ancient Persian and Indian writings may have a common progenitor. And Persian writings also influenced Jewish and subsequently Christian thought. This being said, the above principles and/or practices may still have been developed independently, especially for the Australian aborigines and Native Americans. Their views are likely to be ‘untouched’ by European and Middle Eastern cultures. Furthermore, the influence of European and Middle Eastern thought on Chinese thought and vice versa also seem tenuous. Considering the above, it remains plausible that at least some of these communities developed these principles and/or practices independently from each other. But even if we accept that the shared moral principles and/or practices across cultures have the same origin, this, by itself, does not explain why there is widespread approval of the principles of and/or practices about the golden rule and alleviating suffering. The fact that a principle/practice can be transmitted from one place to another does not explain why a view that

caught on in one place caught on in another place. If I must concede that shared moral principles and/or practices could not have independently developed, I can drop the claim that EAMs can explain how shared moral principles and/or practices across cultures could have independently developed. And adopt the claim that EAMs can explain how shared moral principles and/or practices transmitted from one place to another could have caught on in this other place.

Explaining Moral Unanimity

With what I mean by moral unanimity set out, I can now show how EAMs can explain moral unanimity. I start with a refresher on what unites the various EAMs. In chapters two and three, the reader saw that EAMs come in different shapes and sizes. EAMs differ on what and how emotions (broadly speaking) ground moral judgments, but they are united by the common theme (either expressly or implicitly) that all healthy human beings have some emotions that in one way or another ground some moral judgments. As Hume argues, moral judgment “depends on some internal sense or feeling, which nature has made universal in the whole species” (Hume 1777:1975, 173). For Blackburn, some pro-/con-attitudes ground moral judgment. For Slote, empathic warmth/chill. For Zagzebski, anger and pity can ground moral judgment. And for Mencius, the four emotive sprouts ground moral judgment. Though they differ on what and how emotions ground moral judgments, they (explicitly or implicitly) affirm the universality of their respective grounding emotions. Considering above, the theme underlying this section is that *if all healthy human beings share some emotions that in one way or another ground some moral judgments, this can explain why communities across cultures share some moral principles and/or practices.*

How creatures like us ended up with some shared emotions. A plausible answer to this question is that we share a common evolutionary past. Allow me to explain with reference to Haidt’s moral foundations theory. He argues that there are universal moral “taste receptors” (Haidt 2012, xv). These are cognitive modules that are “adaptations to longstanding threats and opportunities in social life” (Haidt 2012, 144). These modules draw our attention to specific events (e.g. suffering) and can trigger certain emotions (e.g. compassion). The point that I would like to tease out of the moral foundations theory is that certain emotional responses are adaptations in the sense that they are characteristics that improve our chances of survival as social creatures. Here are two examples.

Compassion can draw our attention to and can be triggered by the suffering of others. As Jennifer Goetz et al succinctly argue, “[t]he state like experience of compassion, and the trait like tendency to feel compassion, fall under the purview of three evolutionary arguments: that compassion evolved as part of a caregiving response to the vulnerable offspring, that compassionate individuals were preferred in mate selection processes [such individuals are more likely to offer care and protection], and that compassion emerged as a desirable trait in cooperative relations between non-kin” (Goetz, Keltner and Simon-Thomas 2010, 25). Anger can draw our attention to and can be triggered by an unequal allocation of resources. As Leda Cosmides et al argue, “[t]here is increasing evidence that anger is the expression of a neurocomputational system evolved to bargain for better treatment, by threatening to either inflict harm (punishment) or withdraw cooperation” (Cosmides, Guzmán and Tooby 2019, 205). These emotional responses can ensure the sustainability of social life, which is in turn essential to our survival. If this aspect of the moral foundations theory is correct, what can follow from the above is that humans as social creatures share some universal emotions. If all humans have common ancestors who survived as a result of possessing certain emotional dispositions, it is also plausible that all humans inherited these emotional dispositions. This claim that all humans have some universally shared emotions is further supported by Ekman’s experiments regarding the Fore people. The culturally insulated subjects were able to interpret “faces as showing the same emotions as did the people from the literate cultures”. They were also asked to show how they would look like if they felt certain emotions, the “expressions they showed were once again the common universal expressions [like anger, sadness, happiness and disgust]” (Ekman 1980, 93-94).

How some shared emotions ended up as grounding emotions for some moral judgments. In the preceding paragraph, we saw a plausible explanation on how humans across cultures ended up with some shared emotions. We are now at a point where we can ask how such emotions ended up as grounding emotions for some moral judgments. Darwin provides us with a plausible answer with an account of our transition from social to moral animals; from creatures motivated to act in certain ways to creatures capable of judging in certain ways. Darwin argues that “... the social instincts, — the prime principle of man's moral constitution — with the aid of active intellectual powers and the effects of habit, naturally lead to the golden rule, ‘As ye would that men should do to you, do ye to them likewise;’ and this lies at the foundation of morality” (Darwin 1871: 2009, 106). As we can see here, Darwin argues that social instincts are the starting point toward the capacity to make

certain moral judgments. We are thus presented with a starting point and endpoint. The rest of this section is in part to ‘flesh out’ what is in between the starting point and endpoint. Let us begin.

Darwin argues that “the social instincts lead an animal to... feel a certain amount of sympathy with them, and to perform various services for them” (Darwin 1871: 2009, 72). We see this in nonhuman primates. A chimpanzee mother after hearing the whimpering of her child can proceed to attend to its needs (de Waal 2006, 25). A fellow chimpanzee can express consolation behaviour toward another chimpanzee who lost a fight (de Waal and van Roosmalen 1979). We also see such social instincts at work in humans. Human mothers are anxiously tuned to the whimpering of their own children and would usually attend quickly to their needs. We console each other and empathize with the loss suffered by our family and friends.⁴⁸ The behavioural similarities between humans and nonhuman primates, however, do not show that the latter can make moral judgments. Nonhuman primates may express what appears to be emotions. They may be motivated as such to act in certain ways. But as Hume argues, in such animals, “want of a sufficient degree of reason may hinder them from perceiving the duties and obligations of morality” (Hume 1896, 468). Darwin puts it in another way by arguing that “[a] moral being is one who is capable of comparing his past and future actions or motives, and of approving or disapproving of them. We have no reason to suppose that any of the lower animals have this capacity...” (Darwin 1871: 2009, 88-89). To put this point in the context of outlining how we got from instinctive social behaviour to the making of moral judgments, the possession of social instincts and certain emotions are one step toward being able to make moral judgments.

⁴⁸ At this point, it is important to highlight the danger of anthropomorphism. We do not want to incorrectly attribute human emotions to animal behaviour. This being said, in avoiding anthropomorphism, we would not want to fall into the trap of ignoring the similarities between nonhuman primates and humans either. There are certainly similarities. De Waal describes a very striking example: “... two male chimpanzees who have been chasing each other, barking, and screaming, and afterwards rest in a tree. Ten minutes later, one male holds out his hand, begging the other for an embrace. Within seconds, they hug and kiss, and climb down to the ground together to groom” (de Waal 2014, 55). This behaviour is strikingly similar to reconciliatory behaviour in humans. And if we accept the Darwinian view that the difference between humans and animals (nonhuman primates especially) is a matter of degree and not kind, the modest suggestion that nonhuman primates possess some rudimentary forms of emotion seems plausible. As Hume argues (albeit in a far more ambitious manner): “Tis from the resemblance of the external actions of animals to those we ourselves perform, that we judge their internal likewise to resemble ours; and the same principle of reasoning, carry’d one step farther, will make us conclude that since our internal actions resemble each other, the causes, from which they are deriv’d, must also be resembling. When any hypothesis, therefore, is advanc’d to explain a mental operation, which is common to men and beasts, we must apply the same hypothesis to both” (Hume 1896, 226).

We can now consider what is involved in the 'social instincts' with more detail. Social animals require some level of interaction to respond to the dangers and needs of the group (de Waal 2006, 25). Sensitivity to the emotional expressions and vocalisations of others in the group may have increased the chances of survival for social animals. And insensitivity to such expressions and vocalisations can often be fatal to members of the group. De Waal provides us with an example where "a female chimpanzee lost a succession of infants despite intense positive interest because she was deaf and did not correct positional problems (such as sitting on the infant, or holding it upside-down) in response to its distress calls" (de Waal 2006, 24). This emotional sensitivity can start with the unconscious mimicry of the emotional expressions of others. More specifically, Elaine Hatfield et al refer to this phenomenon as 'emotional contagion', i.e. "the tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person's and, consequently, to converge emotionally" (Hatfield, Cacioppo and Rapson 1993, 96). In our context, we can divide this definition into two parts. The first part involves the automatic mimicry of emotional expressions and vocalisations. In terms of mimicking facial expressions, infants "mimic facial expressions of emotion shortly after birth and continue to do so throughout their lifetimes" (Hatfield, Cacioppo and Rapson 1993, 97). As for mimicking emotional vocalisations, one baby who cries in a maternity ward can set off the cries of other babies. This automatic mimicry also extends to adult humans. Ulf Dimberg et al's experiment is a case in point (Dimberg, Thunberg and Elmehed 2000). His subjects were briefly presented with images of different emotional expressions on a screen. The brevity in which these images were shown did not allow them to be registered in the subjects' consciousness. Despite this, the subjects had corresponding facial expressions to the images. The subjects frowned when shown an angry face. The subjects smiled when shown a happy face. The second part involves the feeling of the emotions associated with the expressions and vocalisations mimicked. As Hatfield et al argue, "[w]hen people produced facial expressions of fear, anger, sadness, or disgust, they were likely to feel the emotion associated with those specific expressions" (Hatfield, Cacioppo and Rapson 1993, 98). Here we see that, by automatically mimicking another's emotional expression, one can 'catch' the emotion of the other. Adam Kramer et al also argue that, "[e]motional states can be transferred to other via emotional contagion, leading people to experience the same emotions without their awareness" (Kramer, Guillory and Hancock 2014, 8788). In an experiment, they sought to test whether the emotional contagion of Facebook users "occurs outside of in-person interaction

between individuals by reducing the amount of emotional content in the News Feed” (Kramer, Guillory and Hancock 2014, 8788). The results show that “people who had positive content reduced in their News Feed, a larger percentage of words in people’s status updates were negative and a smaller percentage were positive. When negativity was reduced, the opposite pattern occurred” (Kramer, Guillory and Hancock 2014, 8789). One key finding of this experiment is that there appears to be no need for nonverbal cues from another to ‘catch’ the emotional state of another. But the point that I wish to emphasize is that de Waal’s, Hatfield and Dimberg et al’s findings on emotional contagion remain relevant to this day. And this ‘catching’ of the emotional states of others paves the way for the development of empathy.

Our emotional contagion may have gradually developed into empathy when we have “an increasing appreciation of the precise circumstances underlying the emotional states of others” (de Waal 2006, 28). De Waal’s Russian doll metaphor can help ‘flesh out’ the transition from emotional contagion to full-blown empathy (de Waal 2006, 37-42). We start off with emotional contagion as the ‘core’ doll. The dolls that ‘build on’ this contagion is cognitive empathy (the ability to “assess the situation and reasons for other’s emotion”) and attribution (the ability to “fully [adopt] the other’s perspective”) (de Waal 2006, 39). The metaphor shows how the level of one’s empathy depends on the complexity of their “cognitive overlays” that allow them to adopt the viewpoint of others (de Waal 2006, 30). Though apes exhibit ‘consolation’ behaviour (reassuring one’s companion), monkeys do not. And humans, by far, have superior cognitive overlays that allow us to take on the perspective of others. This ability to take on another’s perspective is built on our capacity for sympathetic concern, which in turn has its basis in our automatic mimicry of the emotional states of others. There is also more recent evidence supporting this Russian doll metaphor, where our accuracy in identifying another’s emotional expression is reduced if our mimicry of that emotional expression is dampened through Botox injections (Neal and Chartrand 2011).

As we can see with the preceding two paragraphs, emotional contagion and empathy can lead us to feel certain emotions. Applying this point to Slote’s empathy-based sentimentalism, our empathic capabilities allow us to catch the ‘warmth’ of Mother Teresa helping those in need or the ‘chill’ of Hitler in ordering the killing of Jews. And moral judgments in turn have empathic

warmth/chill as referents. And if we apply the point to Zagzebski's and the Mencian account of moral judgments, we can see how our empathic capabilities allow us to feel compassion for another. In Zagzebski's account, empathy (or at least emotional contagion) can be a prerequisite to feeling compassion for the suffering of another. Since for Zagzebski to be in a state of compassion involves seeing someone as pitiable, empathy/emotional contagion can plausibly be a prerequisite to seeing someone as pitiable. Similarly for the Mencian account, empathy/emotional contagion is plausibly involved in the Mencian agent's witnessing of a child who is about to fall into a well. And this empathy/emotional contagion can plausibly be a precursor to the compassion that is felt by the Mencian agent. And since compassion's cognitive component involves the judgment that the child is about to suffer and this anticipated suffering is a reason to act in helping ways, empathy/emotional contagion can plausibly be a precursor to such a judgment. A chimpanzee mother's empathy toward her whimpering child can 'lead her' to feel compassion. Her compassion can motivate her to alleviate the suffering of her child. But given her limited cognitive abilities, she is not able to see such an act as the right thing to do. This is different for human mothers. A human mother's empathy toward her whimpering child can 'lead her' to feel compassion. Her compassion can not only motivate her to alleviate the suffering of her child but also allow her to see such an act as the right thing to do.

Our linguistic capability is plausibly another key factor that helped us transition from creatures motivated to act in certain ways to creatures capable of judging in certain ways. Darwin argues that "after the power of language had been acquired and the wishes of the members of the same community could be distinctly expressed, the common opinion how each member ought to act for the public good, would naturally become to a large extent the guide to action" (Darwin 1871: 2009, 72). Similarly, Philip Kitcher argues that:

"[a]lthough a precursor to full ethical life was available prior to the acquisition of language, explicit consideration of what to do only came into being once our progenitors could speak to one another and to themselves... From fifty thousand years before the present to the late Paleolithic, human beings engaged in numerous 'experiments of living', and the most successful elements of their practices were inherited by those who crafted the rules for the earliest cities, and were passed on to the ethical codes of today" (Kitcher 2014, 248-249).

I draw two points from Darwin's and Kitcher's arguments. The first is that our pro-/con-attitudes toward x can be conveyed to others upon acquiring the power of language. Language allows us to put our attitudes toward actions/persons into words and communicate them to others. The second is that the verbalised attitudes in a community can guide action in the same community. Language allows us to express our attitudes in several ways. A con-attitude toward x can be expressed as ' x is bad' or ' x is wrong'. And as argued in chapter two on Blackburn's expressivism, the utterance that ' x is bad' or ' x is wrong' need not be confined to an expression of a con-attitude toward x . The utterance can be an expression of a cluster of attitudes involving x . It can be an expression of a con-attitude toward x , a con-attitude toward those who commit x , a con-attitude toward those who do not express a con-attitude toward x , and a con-attitude toward those who express a pro-attitude toward x . If so, such attitudes can also be expressed as ' x should be prohibited', 'those who commit x should be punished', 'those who do not condemn x should be punished', etc. Considering this, if we humans share some emotions like compassion toward the suffering and resentment toward those who cause or refuse to alleviate suffering, such attitudes can be expressed in the manner similar to what we see in the *Book of Deuteronomy*: "You shall not oppress a hired servant who is poor and needy, whether he is one of your brethren or one of the sojourners who are in your land within your towns; you shall give him his hire on the day he earns it, before the sun goes down (for he is poor, and sets his heart upon it); lest he cry against you to the LORD, and it be sin in you" (Deuteronomy 24:14-15). With the above, we see how our linguistic capabilities facilitated the transition from creatures motivated to act in certain ways to creatures who judge in certain ways. We start off with certain pro-/con-attitudes toward x . With our linguistic capabilities these attitudes can be verbally expressed, codified into writing and used as means to guide conduct within a particular community. In short, we humans started off with some pro-/con-attitudes and we ended up via our linguistic capabilities with such attitudes being verbalised and codified.

Explaining moral unanimity. At the start of this section, we saw how humans across cultures can end up with an array of shared emotions. We not only saw a plausible explanation of how humans ended up sharing certain emotions (e.g. compassion), we also saw how these emotions individually have (at least at a high level) a set eliciting condition (e.g. suffering). After which, we saw how these shared emotions with such eliciting conditions ended up grounding some moral judgments. The point I would like to emphasize now is that this narrative can plausibly explain some instances of moral unanimity. It paves a way toward making sense of the observation that

humans across cultures share certain moral judgments, rules and practices. We have already seen how some shared emotions can explain the ‘anatomy’ of some moral judgments. Drawing on the water and sprout metaphors in the previous chapter and the moral foundations theory, we saw how humans plausibly have certain shared inherent emotional dispositions to judge in certain ways. In the rest of this section, I intend to show how the pancultural status of some moral principles and/or practices can be explained by clustering these principles/practices around certain emotions. There are likely to be other moral principles/practices that can be clustered around other emotions, but I would limit my discussion to three emotions, i.e. compassion, anger and deferential feelings.

First, we can explain pancultural rules and practices on alleviating suffering by having them clustered around compassion. With the above, I outlined the imperative to care for sojourners, orphans and widows (Deuteronomy) and practices of caring for those in need (Australian aborigines and Native Americans). The following are some other similar moral principles and/or practices across cultures. In ancient Babylon, we see that those who care for the weak are regarded as praiseworthy: “Whoso makes intercession for the weak, well pleasing is this to Samas” (Babylonian. *ERE* v. 445). We see also that those who cause harm are expected to be punished: “Who meditates oppression, his dwelling is overturned.” (Babylonian. *Hymn to Samas. ERE* v. 445). In ancient Egypt, we see a sense of duty toward the poor: “I have given bread to the hungry, water to the thirsty, clothes to the naked, a ferry boat to the boatless” (Ancient Egyptian. *ERE* v. 446). In ancient India, we see the obligation to care for the poor: ‘He who is asked for alms should always give.’ (Hindu. Janet, i. 7).⁴⁹ These principles and/or practices of care can plausibly be clustered around and grounded in the universally shared emotion of compassion. By applying Blackburn’s expressivist account to these points (as an example), if people across cultures have pro-attitudes (via compassion) toward those who alleviate suffering, this can plausibly explain why we universally share principles and/or practices that deem or suggest the alleviation of suffering as praiseworthy or obligatory.

⁴⁹ These passages are directly quoted from C. S. Lewis’ *The Abolition of Man* (C. S. Lewis 1944: 2001, 85, 87, 96, 97).

Second, some pancultural principles and/or practices on justice can also be explained by having them clustered under a specific type of anger. This sort of anger appears to be universal in light of findings on the ultimatum game (Appiah 2008, 132-133). The game involves Persons A and B. Person A is instructed to apportion an amount with Person B. Person B can either accept or reject the apportionment. If Person B accepts the apportionment, both Persons A and B will obtain what was apportioned. But if Person B rejects it, neither will get anything. The rational thing to do for B is to accept the apportionment regardless of how unfair it is. It is better for Person B to get something rather than nothing. However, it is often the case that Person B will reject the unfair apportionment so to punish Person A. Person B appears to be motivated by anger to punish Person A for taking more than what he deserves. As Kwame Appiah argues, “[p]eople everywhere seem to have a natural tendency to punish antisocial behaviour of this sort...” (Appiah 2008, 132). This shared anger at another’s undeserved acquisition of resources can explain why we universally share principles that prohibit what is regarded as illegitimate acquisitions of resources and practices that govern the allocation of resources. By applying the above to Zagzebski’s account of moral judgment (for example), if we universally judge (via anger) an undeserved acquisition of resources as offensive, this can explain why we share the following pancultural principles/practices. In ancient Babylon, we see a rejection of unjust acts: “Has he drawn false boundaries?” (Babylonian. List of Sins. *ERE* v. 446). In ancient Israel, we see the prohibition of theft (which is a kind of injustice): ‘Thou shalt not steal’ (Exodus 20:15). In the Australian Aborigines we see the practice of giving unto each what they deserve: “If the native made a “find” of any kind (e.g. a honey tree) and marked it, it was thereafter safe for him, as far as his own tribesmen were concerned, no matter how long he left it” (Australian Aborigines. *ERE* v. 441).⁵⁰ These principles and/or practices share a common theme that ‘one should be given what one deserves’ or ‘one should not take what one does not deserve’ and can plausibly be explained by our shared anger toward those who take or are given what they do not deserve.

Third, some pancultural principles and/or practices on respecting authority can be explained by clustering them around feelings of deference. Let us begin by what appears to naturally evoke such feelings of deference. Jean Piaget argues that children feel a *primitive* form of respect toward adults, i.e. a “respect felt by the small for the great...” (Piaget 1932, 102). He goes on to argue that

⁵⁰ These passages are directly quoted from C. S. Lewis’ *The Abolition of Man* (C. S. Lewis 1944: 2001, 93, 94)

“[i]t is a fact that the child in the presence of his parents has the spontaneous feeling of something greater than and superior to himself. Thus respect has its roots deep down in certain inborn feelings and is due to a *sui generis* mixture of fear and affection which develops as a function of the child’s relation to his adult environment” (Piaget 1932, 379). What I gather from Piaget’s argument is that: (a) deferential feelings are commonly evoked when we come into the presence of the ‘great’, (b) children generally have such deferential feelings toward their parents (whom they see as ‘great’), and (c) deferential feelings serve as a precursor to a more cognitively loaded emotion of respect. If Piaget is correct, these shared deferential feelings in children and how they serve as a precursor to respect appear to be able to explain why people across cultures seem to believe that parents should be respected. This is evident in the ancient world as well. In ancient China, respect for parents was an integral part of its moral fabric: “A young person should be filial when at home...” and “[m]ight we not say that filial piety... constitute the root of Goodness?” (Confucius c 500 BC: 2003, 1.6, 1.2). In ancient Israel, we see the commandment “[h]onour your father and your mother, that your days may be long in the land which the LORD your God gives you” (Exodus 20:12). We see also how dishonouring one’s parents was dealt with: “For every one who curses his father or his mother shall be put to death; he has cursed his father or his mother, his blood is upon him” (Leviticus 20:9). These principles and/or practices share a common theme that ‘we should respect our parents’ and can plausibly be explained by the shared deferential feelings of children toward their parents. By applying the above to the Mencian account of moral judgment (for example), if we universally judge/perceive (via the cognitive component of respect) parents as worthy of respect and this as a reason for acting in respectful ways, this can explain why we universally share the principles that require us to act in respectful ways toward our parents and practices that punish those who do not.

The reader may object by saying that the principle that ‘we should respect our parents’ is no longer universally accepted. And the Western culture is a case in point. I draw the reader’s attention to Haidt’s anecdote about a Jordanian living in the United States. He wanted to return to Jordan because he did not want his son to say ‘fuck you’ to him (Haidt 2012, 165). Such an objection, however, may be too hasty. Even if it is true that western children are less respectful toward their parents when compared to other cultures, it does not follow from this that the West does not accept the principle that ‘we should respect our parents’. The rampant failure of children in respecting their parents in a particular community can co-exist with the expectation placed on children to

respect their parents. Ubiquitous complaints by western parents on the disrespectfulness of their children reveal at least that they expect their children to respect them. Parents *and* children may also believe that they should respect their parents but admit exceptions to this principle. More specifically, parents and children in the West may believe that parents are owed respect unless they act in certain ways that deprives them of such respect. Absent or abusive parents may no longer deserve respect from their children. With this paragraph, we can see that it at least remains plausible that ‘we should respect our parents’ is universally accepted (even in the West).

In sum, we saw in this section (a) that humans share some universal emotions, (b) a plausible explanation of how humans ended up with these universal emotions, (c) a plausible evolutionary explanation of how these emotions ended up as grounding emotions for moral judgments (the transition from social to moral animals), and (d) how EAMs along with (a) to (c) can explain certain principles and/or practices that enjoy widespread approval across cultures and could have been independently developed. With these I submit that this EAM model can plausibly explain instances of moral unanimity (a salient feature of morality).

Moral Diversity

In the preceding sections, we saw that there are clear instances of moral unanimity. But this is not the only salient feature of morality. There are also clear instances of moral diversity. Allow me to explain. John Finnis argues that:

“All human societies show a concern for the value of human life; in all, self-preservation is generally accepted as a proper motive for action, and in none is the killing of other human beings permitted without some fairly definite justification. All human societies regard the procreation of a new human life as in itself a good thing unless there are special circumstances. No human society fails to restrict sexual activity; in all societies, there is some prohibition of incest, some opposition to boundless promiscuity and to rape, some favour for stability and permanence in sexual relations” (Finnis 2011, 83).

The emphasis of his argument appears to be that there is some level of moral unanimity and (as examples) all communities in one way or another restrict (a) the killing of human beings and (b) sexual activities. However, the unanimity involved is only at a high level. If we probe deeper, we

can see how these prohibitions pan out in very different ways. Epictetus argues that “[t]his is the dispute among the Jews and the Syrians and the Egyptians and the Romans; not whether holiness should be preferred to all things and in all cases should be pursued, but whether it is holy to eat pig’s flesh or not holy” (Epictetus c 108 AD: 1904, 62). Just as there is no disagreement between the Jews, Syrians, Egyptians and Romans that holiness is required, there is no disagreement across and within communities that some types of killing or sexual activities are prohibited. But just as there are disagreements on what counts as holiness, there are disagreements across and within communities on what specifically are the types of killing or sexual activities that are prohibited. The point here is that while there may be instances of moral unanimity in one sense, there are also instances of moral diversity.

Let us consider the diversity involved in the above prohibitions. I start with the diversity of exceptions involved in the ‘no killing’ rule. In some parts of the world, though one is generally obliged to refrain from killing another, honour killings are an exception. Many disagree that this is a legitimate exception. Rashida Manjoo reports that “[m]urder to cleanse family honour is committed with high levels of impunity in many parts of the world. Although honour crimes have mainly occurred in the vast zone spreading from the Sahara to the Himalayas, it also occurs in other regions and countries with migrant communities” (Manjoo 2012, 11-12). The point I wish to emphasize here is that while many communities regard such a practice of ‘cleansing family honour’ to be abominable, the communities that adopt such a practice may not. We can see a more extreme example in Nazi Germany’s propaganda film *The Eternal Jew* where, as Richard Barsam puts it, the film “argues that Jews are criminals; that they have no soul; that they are different in every way; that killing them is not a crime, but a necessity – just as killing rats is a necessity to preserve health and cleanliness” (Barsam 1973: 1992, 205). For some in Nazi Germany, killing Jews is an exception to the ‘no killing’ rule. I for one and most others would think otherwise. While most communities (I hope) would disapprove of Nazi Germany’s state sanctioned practices of treating Jews as racially inferior, at least some communities within Nazi Germany would have approved, encouraged and/or perpetuated such practices.

Now let us continue with the diversity involved in the types of sexual activities permitted. While in most western countries homosexual intercourse is permissible, such an act is punishable by way

of death penalty in countries like Iran, Saudi Arabia and Yemen (ILGA and Carroll 2016). While most of the west have practices condemning those who *denigrate* homosexuals, there are communities that (to the contrary) have practices condemning those who *are* homosexuals. For some, heterosexual intercourse outside of marriage is permissible. But for others, any sexual relationship prior to marriage is impermissible. While some staunch religious communities have practices that condemn and punish those who engage in pre-marital sex, other communities may not. We can see a more extreme example in the Sambia of New Guinea where boys perform ritual oral sex on adult men (Herdt 1987). For the Sambians, such acts associated with paedophilia are permissible. But for most of us, they would be utterly unacceptable. While the Sambians appear to have practices promoting paedophilia, many communities have practices that condemn and punish those engaged in paedophilic activities.

The above is only a brief survey of moral diversity. The aim is to illustrate that, on a closer and more careful observation of the moral terrain, what appears to be the smooth surface of moral unanimity is riddled with pockets of moral diversity. Even as moral unanimity is a salient feature of morality, moral diversity is also a salient feature.

Explaining Moral Diversity

With moral diversity outlined, I can now show how EAMs can plausibly explain moral diversity. In chapters two and three, we saw how anger, compassion, deference, empathic warmth/chill, etc. can be grounding emotions of some moral judgments. Earlier in this chapter we also saw how these grounding emotions can be universally shared. We are now able to consider: (a) how grounding emotions can be tuned, (b) how the tuning of grounding emotions can influence moral judgments that ensue, and (c) how such moral judgments can influence our moral practices. As a starting point, I note that the underlying theme of (a), (b) and (c) is similar to Flanagan's argument on moral personality. Just as our moral personality is the result of the "growth, regulation, moderation, modification, tuning, and suppression of ancient equipment composed of relatively narrow dedicated processors" (Flanagan 2016, 57), our moral *practices* and *judgments* are the result of the 'growth, regulation, moderation, modification, tuning and suppression' of the grounding emotions.

On the point of how grounding emotions can be tuned, I argued in chapter three that Mencian water and sprout metaphors can serve as explanatory tools on how certain practices can influence our emotional dispositions. I also argued that Mencian emotive sprouts are akin to ‘first drafts’ in the sense that they are open to revision. Though these are Mencian claims, they can apply to EAMs in general. I start with Flanagan’s argument on the distinction between first and second natures. He argues that an agent’s first nature is the “initial settings of dispositional foundations” that the agent has and her second nature is the result of her first nature intermingling with her natural and social ecologies (Flanagan 2016, 105). To put this in context, certain emotional dispositions of an agent have an initial set of eliciting conditions or the agent has initial dispositions to see/feel in certain ways. This would be her ‘first nature’. These eliciting conditions and ways of seeing/feeling can change as the agent engages and interacts with others. The result of her engagement and interaction would be the agent’s ‘second nature’. Let us examine more closely this point on first and second natures in light of two grounding emotions: disgust and shame.

Disgust. This emotion can be directed at different objects. It can be directed at contaminants like faeces. Disgust can also be directed at those who cheat on their partners. One distinguishing feature between them is that the former appears to involve far less learning than the latter. It would seem implausible for a toddler to feel disgust directed at cheats. It appears that far more extensive learning is needed for cheats to be co-opted as an object of disgust. Whereas it also appears that far less learning is needed to co-opt putrid meat as an object of disgust. What can follow from this is that we may have started off with the emotional disposition to react with disgust when triggered by putrid meat (first nature), and we subsequently extended the list of triggers to cheats (second nature). We may have started off with seeing putrid meat as disgusting (first nature), but subsequently learned to see cheats as disgusting as well (second nature). As Nussbaum argues, “[d]isgust... begins with a group of core objects, which are seen as contaminants... Disgust, however, soon gets extended to other objects, through a complicated set of connections” (Nussbaum 2004, 93). This extension can be done via (a) contagion (e.g. regarding cheats as being ‘tainted’ by putrid meat, hence disgusting) or (b) comparison (e.g. associating an action of cheats with the disgusting features of putrid meat). While disgust at putrid meat can be regarded as our first nature, disgust at cheats can be our second nature. The ethic of divinity observable in some

cultures (where the self is seen only as a steward and not an owner of its body) (Haidt 2012, 117), can provide a further illustration of how this extension from first to second nature works at a high level. The disgust of agents in a culture that adopts the ethic of divinity can be tuned in a manner that is sensitive to actions that ‘taint’ the body. As “[t]he body is a temple, not a playground” (Haidt 2012, 117), even consensual bodily acts between parties can be regarded as a ‘tainting’ of the body. Under the influence of the ethic of divinity, agents may express disgust toward pre-/extra-marital sex, certain sexual acts, etc. By association or comparison with what is already seen as disgusting, they may also see these sexual acts as disgusting.

Hume provides us with some insight on how this association can happen. He argues that:

“The long and helpless infancy of man requires the combination of parents for the subsistence of their young; and that combination requires the virtue of chastity or fidelity to the marriage bed...”, and “[o]ne becomes contemptible, no less than odious, when he forgets his duty, which, in this particular, he owes to himself as well as to society... The greatest regard, which can be acquired by [women] is derived from their fidelity; and a woman becomes cheap and vulgar, loses her rank, and is exposed to every insult, who is deficient in this particular” (Hume 1777: 1912).

For Hume, acts of infidelity go against the interests of society (i.e. the stable upbringing of children). As such, men and women are expected to be ‘faithful to their marriage beds’. If they are faithful, they are respectable. If they are unfaithful, they are contemptible and denigrated. While Hume did not explicitly argue for this, the underlying theme here appears to be that those who fail to meet certain expectations are demarcated from those who do meet them. Those who fail to meet them are assigned a lower status. And as disgust is an emotion that “assign[s] to lower status those against whom they are directed...” (Miller 1997, 205), we can see how the association between putrid meat and those who are unfaithful can occur. The denigration process can involve associating the unfaithful with what is already disgusting. Just as putrid meat can be considered the ‘lesser’ of fresh meat, those who are unfaithful can also be considered the ‘lesser’ members of society. Just as putrid meat is repulsive, those who are unfaithful are also repulsive. They deserve to be treated like putrid meat. They should be disposed of or at least kept separate from other members of society.

Shame. This emotion can also be directed at different objects. Shame can be directed at one's own helplessness. This same emotion can also be felt when one plagiarises. On the former, Nussbaum argues that "[w]hen an infant realizes that it is dependent on others, and is by this time aware of itself as a definite being who is and ought to be the center of the world, we can therefore expect a primitive and rudimentary emotion of shame to ensue. For shame involves the realization that one is weak and inadequate in some way in which one expects oneself to be adequate. Its reflex is to hide from the eyes of those who will see one's deficiency, to cover it" (Nussbaum 2004, 183). This primitive shame if felt by infants is a far cry from the shame felt when one plagiarises. If we take Nussbaum's argument for granted, one key distinguishing feature is that the former appears to be universal and the latter community specific. Using words/ideas without citing their sources may be permissible in some communities, but impermissible in others. For example, Elizabeth Heitman and Sergio Litewka argue that "[m]any insist that what the US considers plagiarism represents, instead, the writer's expectation that informed readers will be familiar with the original, authoritative source of certain material" (Heitman and Litewka 2011, 106). What can follow from this is that the act of using words/ideas without citing their sources need not be an eliciting condition for shame or seen as shameful in some non-US cultures. And as Nussbaum argues, "societies have a good deal of room to shape the experience differently, both by teaching different views of what is an appropriate occasion for shame and by linking shame differently with other emotions" (Nussbaum 2004, 185). Different communities can as such have and teach different views of what should be seen as shameful. This can explain why plagiarism is an eliciting condition for shame of agents in some communities but not in others or why agents in some communities see plagiarism as shameful and others do not. In light of the above, while shame directed at personal inadequacy/weakness can be an agent's first nature, shame directed at plagiarism can be the agent's second nature.

With the preceding paragraphs of this section, we saw how shared emotions (despite their evolutionary origins) remain malleable. While an agent's engagement and interaction with others can result in the development of 'second nature' emotional dispositions, the reader may wish to probe further into the details on this shift from first to second nature. I intend to provide further details with a careful inspection of the nature of emotions.

As argued in chapter one, I adopt an ecumenical approach toward emotion theories and presuppose that some emotions are not just brute feelings in that they have some cognitive element that can be influenced by thought. In the same chapter, I also expressed my neutrality on what this cognitive element is. This being said, it would be impractical for me to consider how each variety of what emotion theorists would consider as the cognitive element can explain the malleability of some emotions. As such, I will for the rest of this section focus only on one variety, i.e. Zagzebski's emotion theory. In chapter one, I outlined her theory in detail. For the reader's convenience, I briefly reiterate her position here. For Zagzebski, an emotion is a 'unitary state' in that it is constituted by cognitive and affective elements that cannot be pried apart. Being in an emotional state involves both experiencing a distinctive feeling (affective element) and the perception of an object as possessing a 'thick affective property' (cognitive element). Each emotion has a corresponding 'thick affective concept'. To feel anger for example is to see x as coming under the 'thick affective concept' of being offensive. It is at this stage that we can see how inherent emotional dispositions can be malleable. There appears to be an opening for the practices of a community to influence whether an object comes under a thick affective concept. Allow me to probe into this further with the shared emotion of anger as an example.

Harry Frankfurt argues that "[a]ttributing moral blame is distinctively a way of being angry at the wrongdoer... What makes moral anger understandable and appropriate is that the transgression of an immoral agent consists in his willfully rejecting and impeding the realization of our moral ideal. In other words, he deliberately injures something that we love. That is enough to make anyone angry" (Frankfurt 2006, 47-48). Here Frankfurt appears to put forward three claims. First, anger at x can involve attributing moral blame to x . Second, such anger is understandable/appropriate if x rejects/impedes our moral ideal. Third, our anger at x can be triggered by x 's act of injuring what we love. While all three claims are interesting in their own right, for the purposes of my thesis, I focus on the first and third claims. The first claim is relevant as it confines our discussion to anger as a grounding emotion of moral judgments. As such, anger directed at a window that refuses to open for example would not be the type of anger subject to our discussion. The third claim is relevant as it allows us to see how anger (even if it has evolutionary origins) can be influenced by the practices of a community. Explicit and implicit

teaching within communities on what we should hold dear and what we should be angry about. Practices within communities can influence the object(s) of what we hold dear and our anger. In other words, these practices can tune the eliciting conditions of an agent's anger and what she would see as offensive.

The reader at this point may ask two questions. First, how such practices could have arisen in the first place. Second, why are emotions the kinds of things susceptible to tuning. In terms of the first question, it will be answered in more detail later in this chapter. It would suffice to note for now and at a high level that at least some of these practices could have arisen from emotions in the first place. And the practices that arose out of emotions can, in turn, tune the emotions. As for the second question, Zagzebski argues that emotions are "ways of affectively perceiving the world around us that have conceptual constituents of a distinctive kind" (Zagzebski 2004, 52). In other words, emotions involve an agent's seeing of x (an object/state of affairs in the world) as coming under a thick affective concept. In the context of anger, it involves an agent's seeing of x as offensive. This leaves room for members within communities to reflect and debate on whether x comes under the thick affective concept of the 'offensive'. For Zagzebski, our emotions can "change under the influence of emotions of admired others" (Zagzebski 2004, 52). An agent can slowly cease to see x as offensive (for example) if the person whom she admires does not see x as offensive. The agent can also reflect on whether the emotions of the person whom she admires is worth imitating. The more general point here is that, as some emotions possess a cognitive element that can be influenced by thought, we can see why these emotions are susceptible to tuning.

With the above in mind, the ethic of autonomy and community (Haidt 2012, 112-130) observable in different communities can help illustrate how communities can tune the shared emotion of anger in different ways. Communities have different visions of who we are as a people. We may adopt the 'ethic of autonomy' and have practices that involve the seeing of the self as an autonomous individual and prioritizing its needs and preferences. This may be what we love, our moral ideal. The practices associated with the ethic of autonomy could have arisen from various emotions. For example, the practice of criticizing 'big government' by Republicans in the United States could have arisen from their seeing of excessive governmental intrusion into the private lives of citizens as offensive. This anger directed at 'big government' can have its roots in a more

primitive form of anger, i.e. anger at those who take what is mine. What is mine in the Republican example is not a tangible object but one's freedom. An agent's disposition to anger, as a result, can be tuned in a manner sensitive to violations of autonomy. So tuned, her anger may be triggered by authoritarian governments in Singapore and China. She may see various manifestations of authoritarianism as offensive. Others may adopt the ethic of community and have practices that involve the seeing of the self in a less atomistic way and prioritizing the needs and preferences of the community over self. In such communities, the community itself is regarded as a "real" entity and "must be protected" (Haidt 2012, 116). This may be what they love, their moral ideal. Again, the practices associated with the ethic of community could have arisen from various emotions. For example, the practice of criticizing those who rely on handouts without doing their fair share of work in the community could have arisen out of the seeing of those who do not pull their weight as offensive. This anger directed at free-riders can, again, find its roots in a more primitive form of anger, i.e. anger at those who take what they do not deserve. In such an instance, and though anger is shared across communities, an agent's disposition to anger can be tuned in a manner that is particularly sensitive to threats against social cohesion. So tuned, her anger may be triggered by acts of civil disobedience. She may see various acts of civil disobedience as offensive.

In the preceding paragraph, we saw how, at a high level, practices in different communities can tune the shared emotion of anger. Let us now consider a more specific example, i.e. the nexus between anger and honour killings. I refer again to Manjoo's report on honour killings in the preceding section and tease out two points: (a) in some communities murder in the name of honour is committed against women, and (b) such murders can be motivated by a desire to cleanse the family from the shame brought upon by a daughter or sister. Now consider these points in light of Frankfurt's argument above. People in such communities may have been inculcated with some moral ideal of chastity. Children may have been taught that pre-marital sex is wrong, that such an act violates the 'honour' of their family, that they have to keep themselves 'pure' until marriage. This form of education can involve emotional training. To ensure that children do not violate the 'pre-marital sex is wrong' rule, inculcation techniques such as power assertion, love withdrawal, and induction may have been used by their parents or other family members (Eisenberg 2000). In terms of power assertion, parents may threaten punishment if children violate the rule. Parents may tell their children that they would get very angry and spell out the consequences if they violate the rule. In terms of love withdrawal, parents may withdraw from showing signs of

affection for their children if they find out that their children violated the rule. Children may be ostracised from their families by their parents. And as for induction, parents may call to their children's attention the harm that they would bring to their family if they violate the rule. The children's family would have to endure the shame brought on them. As pre-marital sex is an affront to the ideal of chastity, agents in this communities may consider that the daughter or sister who commits such an act should feel ashamed. And the father or brother should also feel ashamed of her 'impious' act as she is part of his family. As Frazana Bari commented in an interview with Robert Fisk, "[h]onour' for men is connected with women's behaviour because they are seen as the property of the family – and of the community... They have no independent identities, they are not independent human beings. Men also think of women as an extension of themselves. When women violate these standards, this is a direct blow to the man's sense of identity" (Fisk 2010). The father/brother in such communities can as a result of this inculcation be disposed to feel anger at his daughter/sister's violation of a cherished ideal and the shame she brought on the family. The father/brother can be disposed to see the daughter/sister's pre-marital sexual act and the shame she brought to her family as a result as coming under the thick affective concept of the offensive. But for many fathers/brothers in other communities (or even within such communities), they would not see such an act by their daughter/sister as offensive let alone be motivated to kill their daughters/sisters. Perhaps for them, chastity is not regarded as a moral ideal. Regardless, with the above, we see that community specific practices that centre around a moral ideal can play a role in determining whether an agent sees an act as coming under the thick affective concept of the offensive.

Tuning the Emotions

While the reader can see (at a high level) in the preceding section that practices can tune emotions, I have yet to fully explain the mechanics of this tuning process. For the bulk of this section, I intend to do so by setting out a model of how practices can tune emotions. Prior to doing so, I will first set out what I mean by 'practices'.

Practices/conventions. Thus far, we considered the diversity of moral practices and the rules/principles that can come with these practices.⁵¹ Before going any further to consider how *practices* can tune emotions, I set out what I mean by practices or what Ruth Millikan refers to as ‘conventions’ as *patterns of behaviour reproduced under the weight of precedent* (Millikan 2005).⁵² Allow me to explain with more detail. Millikan argues that conventions are in part patterns of behaviour that are “[b]eing perpetuated by reproduction... [They] are handed down” (Millikan 2005, 3). Conventions can be handed down by “being copied from one another directly, or one person may tell another how a pattern goes” (Millikan 2005, 4). Honour killings can in this sense be conventions. Such patterns of behaviour can be ‘handed down’ in the sense that fathers and brothers may have copied their forefathers in killing their daughters and sisters to cleanse family honour. Or fathers and brothers may have been told by their forefathers to kill their daughters and sisters who commit adultery.⁵³ Apart from being reproduced, Millikan argues that “a pattern is considered conventional only if thought to have little tendency to emerge or reemerge in the absence of precedent” (Millikan 2005, 7). Honour killings fit the bill again. Such patterns of behaviour can be perpetuated largely under the “weight of precedent” (Millikan 2005, 7). Fathers and brothers may ‘reproduce’ the killing of their daughters and sisters out of blind conformity to a cultural tradition, out of respect for this tradition, out of fear of what other members of the community would think of them if they do not conform to tradition, etc.

Millikan’s account of conventions also has a broader scope when compared to David Lewis’ account of conventions as solutions to coordination problems (D. Lewis 1969: 2002, 36). Lewis provides several examples of such conventions. I will consider one:

“In... Oberlin, Ohio, all local telephone calls were cut off without warning after three minutes. Soon after the practice had begun, a convention grew up among Oberlin residents that when a call was cut off the original caller would call back while the called party waited. Residents usually conformed to this regularity in the expectation of conformity by the other party to the call. In this way calls were easily restored, to the advantage of all

⁵¹ For example, we considered the practice of ‘honour killings’ where the underlying rules include ‘thou shalt not commit adultery’.

⁵² For convenience, I would use the term ‘conventions’ for the remaining parts of this chapter.

⁵³ There are other ways in which, for Millikan, conventions can be reproduced. But for the purposes of this section. These examples would suffice.

concerned. New residents were told about the convention or learned it through experience”
(D. Lewis 1969: 2002, 43)

Here we see a coordination problem. Oberlin residents want to communicate beyond three minutes. But their calls get cut off after three minutes. These residents would have to find a workable set of actions on both ends of the call. To solve this problem, a pattern developed. The original caller would return the call after being cut off and the other caller would wait for her call. This is a Lewisian convention as it is a pattern that provides a solution to a coordination problem. Though Millikan’s account would regard such a pattern as a convention, it represents only a subclass of Millikanian conventions. The convention of honour killings (for example) does not appear to be a pattern of behaviour that serves as a solution to a coordination problem. This being said, I accept (as Graham Oddie argues) that “there are a host of coordination problems, or approximate coordination problems which moral [conventions] do seem to address” (Oddie 1999, 263). For example, the Australian aborigines convention (discussed above) of marking of a ‘find’ to indicate that the ‘find’ belongs to a person could have arisen as a solution to the coordination problem of coming to a workable set of actions in light of the scarcity of resources. As we can see, my point here is not to argue that Millikan’s account is better than Lewis’. Rather, I briefly distinguished Millikan’s account from Lewis’ to highlight how the former may be more suited for my discussion in the rest of this section. By adopting Millikan’s account, I do not have to limit my discussion to how *Lewisian* conventions can tune a specific emotion. This is a more prudent approach as I do not need to cautiously filter out patterns of behaviour that are not solutions to coordination problems. For example, the convention of honour killings may well turn out to be a convention that solves a coordination problem, fathers may need to find ways to preserve their daughters’ honour to keep them marriageable. However, it may well not be (see discussion in the preceding paragraph). But regardless of whether this convention turns out to be a solution to the coordination problem of coming to a workable set of actions in regulating inter-family and/or sexual relations, such patterns of behaviour can still count as Millikanian conventions.

Conventional tuning of compassion. Let us start with the shared emotion of compassion. Edmund Burke argues that, “[t]o be attached to the subdivision, to love the little platoon we belong to in society, is the first principle (the germ as it were) of public affections. It is the first link in the series by which we proceed towards a love to our country, and to mankind” (Burke 1909-14, Para. 75).

His argument appears to refer to how moral *behaviour* starts from our care for those who are near and dear to the entire human race. But Burke's point of expanding one's circle of care can also apply to how compassion can be tuned by certain conventions. Allow me to explain.

As argued in chapters two, three and in the sections on moral unanimity above, compassion as a shared emotion can ground moral judgments. And this in turn can explain the pancultural status of some conventions (and rules/principles that come with them) on alleviating suffering. In this sense, we can say that such conventions 'arose' out of compassion. For example, a convention of caring for the poor in our neighbourhood could have arisen out of compassion, out of seeing them and their frailty due to a lack of food (for example) as pitiable. There may also be other conventions that arose similarly out of compassion; caring for the elderly, the weak, the disabled, the orphaned, the stranger, etc. These latter conventions can in turn tune an agent's disposition to feel compassion by adding new eliciting conditions. In this sense, such conventions can expand an agent's circle of compassion. An agent may have started off with an emotion that has fewer eliciting conditions or with seeing fewer objects as coming under the thick affective concept of the pitiable. But in virtue of these conventions, this list can be expanded or the agent can see more objects as pitiable.

Conventions on alleviating suffering need not stop at humans. As Millikan argues, there is a sense in which conventions are arbitrary: "Conventional patterns are patterns for which other patterns, given different historical accidents, might as well have been substituted" (Millikan 2005, 7).⁵⁴ In other words, conventional patterns of behaviour (under different circumstances) could have differed from what they are now in a particular community. What can also follow from this is that while all communities may have conventions on alleviating suffering, some communities (under different circumstances) may have *developed* conventions that involve the alleviation of suffering of not only humans but also animals.

⁵⁴ Lewis also makes the claim that conventions are arbitrary (D. Lewis 1969: 2002, 70)

As an example of how such conventions might have developed, I am reminded of a PETA advertisement stating: 'To animals, all people are Nazis'. The advertisement compares the cruelty of Nazi concentration camps to factory farming. It is aimed at highlighting the inconsistent positions that we have toward concentration camps and factory farming. If we feel compassion for those who suffered in Auschwitz, we should also feel compassion for factory farmed animals. We can also tease out an implicit argument from the advertisement. In standard form, it can look like this:

P1 The Nazis cruelly took innocent lives.

P2 The cruel taking of innocent lives is wrong.

P3 Factory farming cruelly takes innocent lives.

C Factory farming is wrong.

While the aim of PETA may be to present a logical argument against factory farming. The activist group may also have in mind the aim to excite and enhance our natural disposition for compassion toward those who exhibit suffering. The PETA advertisement is reminiscent of Mencius' advice to King Xuan in 1A7 discussed at length in chapter three. In response to the king's question *do I have what it takes to care for my people*, Mencius reminds the king of his compassion toward the ox led to slaughter and argues that with this he has what it takes to care for his people. All the king has to do is to extend his compassion toward the ox to his people. Similarly, the PETA advertisement draws our attention to the compassion we presumably have for Holocaust victims and their relevant similarities with factory farmed animals. The advertisement aims to excite our disposition to feel compassion for Holocaust victims and extend it to factory farmed animals. If this advertisement and other similar ones are successful, a convention of caring for factory farmed animals could have developed.⁵⁵ Such a convention can also be said to have arisen out of

⁵⁵ It is worth noting that while rational argument can play a role in generating a new convention, emotion still plays a crucial role in *generating* and *sustaining* this new convention. The convention is still emotion-based in the sense that it is generated (in part) by adding factory farmed animals as an eliciting condition of compassion. This 'addition' can be explained in the following way. If emotions are concern-based construals, an agent's compassion toward factory-farmed animals would "consist in perceptually assigning roles to aspects of the situation, and thus arranging them in the conceptual order of the defining proposition of..." compassion: it is important for factory farming animals to be flourishing, but such animals are in distress; may their distress be relieved (Roberts 2013, 47). PETA advocates encourage such a construal with words like 'the Holocaust in on your plate', thus 'priming' an agent (i.e. making the agent more prone) to see factory-farmed animals as pitiable. And once the new convention is generated, it is unlikely to be sustained if members of the community do not continue to see factory farmed animals as pitiable.

compassion in the sense described above. This convention could in turn tune other agents' eliciting conditions of compassion or what other agents see as coming under compassion's corresponding thick affective concept.

Conventional tuning of disgust. Let us continue with a more detailed example using the shared emotion of disgust. I turn in particular to how we may emotionally respond to durians. In general, Westerners exhibit strong disgust toward the fruit. But Southeast Asians love it. They not only show a lack of disgust toward durians, they enthrone it as the 'king of fruits'. One question here is why is there such a difference? One answer to this is that conventions may have added/subtracted durians as an eliciting condition of disgust or as objects that come under the thick affective concept of the disgusting. Let us consider first how durians can be added as an eliciting condition of an agent's disgust. As argued earlier in this chapter, an agent's disgust toward contaminants can be extended to other objects by way of comparison. Certain features of durians can be compared with features of faeces. The stench, shape and texture of the fruit can easily be compared with the stench, shape and texture of faeces. This connection once made can allow us to have durians as an eliciting condition of disgust or to see durians as disgusting. In other words, durians can be added as an eliciting condition of disgust by the agent's focus on certain 'hellish' features of durians and/or construal of certain features of durians as 'hellish'. Now let us consider how durians can be subtracted as an eliciting condition. Disgust toward durians can be 'weeded out' by disassociating durians from other objects seen as disgusting. Yes, durians may smell like faeces. But they are also (to an extent) sweet. This allows us to disassociate durians from what we regard as disgusting and associate them with mildly sweet fruits (e.g. guava or papaya). The texture of durians is custardy/creamy. This again allows us to associate them (to an extent) with the texture of (for example) yogurt and honey. Disgust toward durians can be weeded out by the agent's emphasis on certain 'redeeming' features of durians and/or construal of certain features of durians as redeeming.

The key point I would like to tease out of the above is that conventions can play a role in retaining/subtracting durians as an eliciting condition. Given the stench of durians, the connection between their stench and the stench of other contaminants can be easily established. As a result, durians can easily be co-opted as an object of disgust. At this stage, conventions can reinforce this

seeing of durians as disgusting. Many in Western countries find durians disgusting and express their disgust in myriad ways. The stench and taste of durians have been compared with “[F]rench kissing your dead grandmother” or “pig shit, turpentine and onions garnished with a dirty gym sock” (Tan 2014). A convention of ‘criticizing durians whenever the occasion arises’ could have developed out of such expressions. Such a convention, arising out of disgust, can reinforce our disgust toward durians by creating a kind of feedback loop. Our disgust toward durians and subsequent expression of it may have given rise to the convention. The convention can in turn reinforce and amplify our disgust toward durians. In this sense, we can say that conventions that arose out of disgust can tune disgust. Conversely, different conventions could have arisen in different communities. As argued earlier, there is a sense of arbitrariness in conventions. In South East Asia, where many find durians delicious, subsequently express this, pay lots of money to eat them, etc., a convention of ‘praising durians whenever the occasion arises’ could have developed. Such a convention could ‘weed out’ durians as an eliciting condition of an agent’s disgust. As a result of frequent encounters with this convention, an agent who once saw durians as disgusting may gradually cease to see durians as disgusting. What is important to note here is that this convention can also be plausibly understood as arising out of emotions. The joy of eating durians can underlie and motivate our expressions that durians are delicious and/or our paying of lots of money to buy them. In this sense, the convention of durian praising could have arisen out of joy. If so, we can also say that conventions that arose out of other emotions can tune disgust.

With the above in mind, let us consider how conventions can tune *moral* disgust. There are some who are disgusted by homosexuals.⁵⁶ Nussbaum argues that “what inspires disgust is typically the male thought of the male homosexual, imagined as anally penetrable. The idea of semen and feces mixing together inside the body of a male in the neighbourhood inspires the thought that one might oneself lose one’s clean safeness, become the receptacle for those animal products” (Nussbaum 2004, 113). The point to emphasize here is that for Nussbaum disgust at homosexuals involves thoughts on homosexual intercourse where semen and faeces may come into contact in a person’s body. If this is correct, we can start to see how an agent’s initial disgust toward (potential) contaminants like faeces or slimy objects can subsequently include homosexuals as an eliciting condition. Some may see homosexuals as coming under the thick affective concept of the

⁵⁶ It may also be worth noting that there is empirical evidence that negative attitudes toward homosexuals are commonly associated with the emotion of disgust (Herek 1993, Morrison, et al. 2019).

disgusting as homosexual intercourse is likely to involve the interaction of faeces (a contaminant) and semen (what looks like a contaminant). This disgust can give rise to conventions that result in establishing, retaining, and/or reinforcing the link between homosexuals and disgusting contaminants. A convention of 'shunning anyone who professes to be homosexual' could have arisen out of disgust. One such convention underlie the now defunct US military policy 'don't ask, don't tell'. This policy did not prevent closeted homosexuals from being part of the military. But those who are open about their homosexuality can be barred from entry or subject to discharge. The pattern of behaviour underlying this policy can be the regular shunning, belittling, teasing of those who are openly homosexual. Such and other similar patterns of behaviour can in turn reinforce the agent's disposition to see homosexuals as disgusting. Conversely, the agent's 'linking' of homosexuals to what is disgusting can be weakened by other conventions. The University of Auckland for example adopts a policy of 'zero tolerance for discrimination'. Posters, stickers, and pull up banners like the following can be found at various parts of the university premises:

Our University

- ✓ safe
- ✓ inclusive
- ✓ equitable

~~racism
sexism
ableism
ageism
homophobia
transphobia~~

**ZERO tolerance
for discrimination**

He wāhi whakatoihara kore



www.equity.auckland.ac.nz/zerotolerance

The overarching aim of putting up such messages is in part to “promote a positive culture which celebrate differences, challenges prejudice and ensures fairness” (University of Auckland 2019). This serves as a backdrop to a convention where students or staff who express their disgust at homosexuals would be criticized and/or disciplined. Students or staff so criticized or disciplined can start to disassociate homosexuals from what is disgusting. Such an agent may gradually come to see her disgust at homosexuals as prejudicial. The agent may realize that this can make

homosexuals feel unwelcome, inferior, belittled, etc. Such thoughts can influence the agent's disgust in the sense that she may gradually weaken the link between homosexuals and disgust. She may eventually no longer see homosexuals as disgusting or have homosexuals as an eliciting condition of her disgust.

At this juncture, I present some empirical evidence to support the claim that conventions can weaken the link between disgust and homosexuality. I refer in particular to a form of emotional regulation, known as cognitive reappraisal (Gross and John 2003, Feinberg, et al. 2014). This reappraisal involves "rethinking the meaning of affectively charged stimuli or events in terms that alter their emotional impact" (Ochsner and Gross 2008, 154). In an experiment by Feinberg et al, subjects were recruited to watch three film clips, one of which from *Brokeback Mountain*, showing two men kissing. The control group were instructed that after watching the clips they would "answer various questions regarding the clips". Other subjects were instructed to engage in cognitive reappraisal, that they should "try to think about what [they] were seeing in such a way that [they] don't feel anything at all" (Feinberg, et al. 2014, 518). Both groups were subsequently asked questions on whether homosexual relationships are immoral and whether they support same-sex marriages. As it turns out, the "[c]onservatives in the reappraisal condition were more accepting of homosexual relationships relative to their counterparts in the control condition" (Feinberg, et al. 2014, 519). The experiment suggests that an agent who consciously chooses to unemotionally appraise what may be a disgust eliciting event (e.g. a homosexual activity) may no longer see the event as wrong/immoral. In our context, as the judgment 'homosexual acts are wrong/immoral' can be an expression of the emotional state of seeing homosexual acts as disgusting (see chapter two), cognitive reappraisal opens up the possibility of removing homosexual acts as an object of disgust. And such cognitive reappraisal may be prompted by a convention where students or staff who express their disgust at homosexuals would be criticized and/or disciplined.

At this point, we can also see that patterns of expectations can come with patterns of behaviour (conventions). In this instance, we see that the expectation that agents should not express disgust at homosexuals (rule/principle) can come with the pattern of sanctioning agents who express disgust at homosexuals (convention). Bruno Verbeek argues that "a convention is a stable pattern

of interdependent *expectations of behavior*. The existence of such a pattern is a rule of conduct, a norm, in the group in question [emphasis added]" (Verbeek 2008, 80). While Millikanian conventions are identified as patterns of *behaviour*, we see that Verbeekian conventions are identified as patterns of *expectations*. The point I would like to emphasize here is that while I have not adopted Verbeek's account of conventions, his account does shed light on the expectations (rules/principles) that can come with Millikanian conventions. Furthermore, his account "calls attention to the mental states of the agents conforming to a convention..." (Verbeek 2008, 77). While the psychological state referred to by Verbeek is that of a state of expectation, the psychological states I refer to are emotional states. For Verbeek, patterns of expectations can come with patterns of behaviour. I agree with Verbeek on this point but take a step further by claiming that emotional states can underlie patterns of behaviour and expectations. Emotions can underlie the convention of sanctioning those who express disgust toward homosexuals and the expectations that come with this convention. Anger can underlie the convention/expectation in the sense that the former (the seeing of expressions of disgust toward homosexuals as offensive) sustains the latter. Compassion can also underlie the convention/expectation. The latter can be sustained by the seeing of the maltreatment of homosexuals as pitiable. The convention/expectation sustained by anger and compassion can in turn tune disgust.

Conventional tuning of more than one emotion. In addition, the same set of conventions can tune more than one emotion involving the same object. More specifically, the same convention can add x as an eliciting condition of emotion A, while subtracting x as an eliciting condition of emotion B. Allow me to use as a hypothetical example a Nazi who sends internees to gas chambers to illustrate my point. Let us assume that this Nazi while sending an internee to the gas chamber felt no compassion for the internee's trembling but instead felt disgust at the internee. Here the internee's trembling is like the trembling of the ox led to slaughter in 1A7 of the *Mencius*. But unlike King Xuan who felt compassion for the ox, the Nazi felt no compassion for the internee. Instead, the Nazi felt disgusted by the internee. A question here is how can we explain the coming about of this Nazi's phenomenological experience. Let us begin with his disgust. His disgust may have resulted from social engineering. Nussbaum argues that:

"The stock image of the Jew, in antisemitic propaganda, was that of a being disgustingly soft and porous, receptive of fluid and sticky, womanlike in its oozy sliminess, a foul

parasite inside the clean body of the German male self. When Jews were depicted in fairy tales for children, they were standardly represented as disgusting animals who had these same properties. Thus for Hitler (and not only for him), the Jew is a maggot in a festering abscess, hidden away inside the apparently clean and healthy body of the nation” (Nussbaum 2004, 108).

If Nussbaum is correct, Nazi propaganda that compared Jews to what the agent already sees as disgusting can result in the agent’s seeing of Jews as disgusting. A convention of treating Jews as racially inferior could have arisen out of this disgust. This convention in turn can reinforce the agent’s disgust toward the Jews. Similarly, the Nazi guard’s disgust toward the internee could have resulted from him being subject to the influence of anti-Semitic propaganda and the convention that could have arisen out of it. As for the Nazi guard’s lack of compassion, the same convention could have subtracted ‘the Jews’ as an eliciting condition of his compassion. The convention can turn the Nazi guard’s attention away from what one would usually see as features of the thick affective concept of the pitiable (e.g. trembling resulting from an impending death) to the disgusting features ‘superimposed’ on the Jews. In sum, the same convention that plausibly arose out of disgust can reinforce the seeing of x as disgusting while weakening the seeing of the maltreatment of x as pitiable.

Emotional scripting. With the above, we see that (a) some conventions could have arisen out of compassion, anger and/or disgust, and (b) these conventions can in turn tune an agent’s disposition to feel compassion and/or disgust. For the rest of this section, I intend to show how emotional scripts are involved in this conventional tuning process. Cheshire Calhoun argues that “[e]motional scripts specify the proper object of the emotion (such as the loss of a parent or an insult by an inferior), the physical symptoms (such as weeping or sighing), the goal-directed behavior (such as seeking revenge), and the fantasies or patterns of thought (such as imagining kisses or obsession with signs of betrayal) that typify an emotion” (Calhoun 2001, 220). For Calhoun, scripts specify these four aspects. But I will focus on how scripts specify the object(s) of an emotion for the purposes of this section. We can also understand such scripts as norms that regulate emotions. Jon Elster argues that “[t]here are normative expectations about which emotions one should *express* under specific circumstances, and even norms regulating what emotions one is expected to *feel*...” (Elster 1999, 156-157). Here Elster’s argument appears to be

that scripts specify when and/or at what object(s) *should* an emotion be expressed/felt.⁵⁷ Regarding scripts that specify when and/or at what object(s) should an emotion be *expressed*, Ekman argues that there are *display* rules specifying “who can show what emotion to whom, when” (Ekman 1980, 87). As for scripts that specify when and/or at what object(s) should an emotion be *felt*, Arlie Hochschild argues that there are *feeling* rules that can be made manifest by “the way others react to what they think we are feeling” (Hochschild 1983, 58). While Calhoun, Elster, Hochschild and Ekman may differ on what emotional scripts regulate (i.e. emotional expression and/or experience) and what these scripts are exactly about, they are united at least at a high level that the scripts regulate what should be an emotion’s object(s). Prima facie, the existence of such scripts is not only uncontroversial but also ubiquitous. We are likely to have encountered in our everyday lives remarks like ‘you should (not) feel ashamed of what you have done’, ‘you should (not) feel compassion for that person or that group of people’, ‘you should (not) be angry at her’, etc. These remarks can be understood as scripts in that they provide information on what shame, compassion, anger should (not) be directed at. If an agent internalizes these scripts, her disposition to feel shame, compassion and anger at certain objects can be tuned accordingly.

With the preceding paragraph in mind, we can also see that there can be at least two kinds of expectations that can come with conventions. The first would be expectations of conduct. ‘You should *kill* your adulterous daughter/sister’ can (for example) be the expectation of conduct that comes with the honour killing convention. The second would be expectations of feeling. ‘You should feel anger/shame at your adulterous daughter/sister’ can be the expectation of feeling that comes with the same convention. Emotional scripts can count as such expectations of feelings. I will use examples already discussed in this section to further illustrate my point. On compassion, scripts like ‘an agent should feel compassion toward certain groups of people/animals’ could come with the conventions/expectations relating to the care of the elderly, the weak, the poor, the disabled, the orphaned, the stranger, factory-farmed animals, etc. On disgust, scripts like an agent should (not) feel disgust toward those who are open about their homosexuality could have come with a convention that condemns anyone who professes to be homosexual or a convention where those who express their disgust at homosexuals would be criticized and/or disciplined. We can

⁵⁷ To avoid doubt, this claim does not necessarily entail the claim that the normativity of emotions is derived from the normativity of convention-defined scripts. While a script may specify when and/or at what object(s) should an emotion be expressed/felt, such a script could have, like conventions, arisen out of emotions. The script ‘you should be angry at *x*’ could have arisen out of an agent’s initial anger at *x*.

see a more concrete example of emotional scripts and how these scripts could have come with conventions in *The Eternal Jew* discussed earlier in this chapter. David Livingstone Smith argues that this film “emphasizes the connection between Jews and filth, decay, and disease in every sector of cultural life” (D. L. Smith 2011, 137). The narrator in the film for example explains that:

“Wherever rats appear they bring ruin, by destroying mankind’s goods and foodstuffs. In this way, they spread disease, plague, leprosy, typhoid fever, cholera, dysentery, and so on. They are cunning, cowardly, and cruel, and are found mostly in large packs. Among the animals, they represent the rudiment of an insidious and underground destruction, just like the Jews among human beings” (D. L. Smith 2011, 137).

The narration appears to describe how the Jews are like rats. But as the film was a Nazi propaganda tool, it is plausible to understand this narration as an attempt to create a script that specifies the Jews as a proper object of disgust. The film can imply that ‘you should feel disgusted by the Jews just as you feel disgusted by rats’. As a result, this script can tune our disgust. As Glen Pettigrove and Nigel Parsons argue, “[t]he vividness with which some emotional scripts are presented in literature, ritual, the popular media, political monuments, and educational curricula will influence... the likelihood of a particular emotional response to certain kinds of stimuli...” (Pettigrove and Parsons 2012, 510). The key point to note here is that the script described above could have plausibly come with the convention of treating Jews as racially inferior. For example, Jews were segregated from the general German populace in pre-war Nazi Germany. They were barred entry to public schools, universities and cinemas in the name ‘Aryanization’ (United States Holocaust Memorial Museum). While some may bar entry unreflectively, others may do so out of the conviction that the Aryan race is ‘pure’ and Jews ‘impure’. The latter can be the ones responsible for the proliferation of emotional scripts like ‘you should be disgusted by the Jews’. Both script and convention could have arisen out of *The Eternal Jew*, which tapped into the existing prejudice of some Europeans toward Jews.

Taking stock. Earlier in this section, I argued that (a) emotions can underlie conventions, and (b) conventions can tune emotions. I can now further refine these claims in the following manner: (a) emotions can underlie conventions *and the expectations that come with the conventions*, and (b) conventions can tune emotions *via expectations of feeling (emotional scripts) that come with the conventions*.

Conflicting conventions. At this juncture, it is apt to point out that a community can have conflicting conventions. This is especially so when a community reaches a certain population. In such a circumstance, there is bound to be some level of heterogeneity in a community's conventions. For example, a section of the community can have a convention of condemning anyone who professes to be a homosexual and another section of the same community can have a convention where those who express their disgust at homosexuals would be criticized and/or disciplined. What can follow from this is that there can be conflicting expectations of feeling (emotional scripts) for any given emotion within a community. In one section of the community, there may be a script that prescribes homosexuals as the proper object of disgust. In another, there may be a script that prescribes homosexuals as an improper object of disgust. On a related point, different scripts may apply to different people within a community due to their difference in status. Hochschild argues that "when one person has higher status than another, it becomes acceptable to both parties for the bottom dog to contribute more. Indeed, to have higher status is to have a stronger claim to... emotional rewards" (Hochschild 1983, 84). For example, the poor (and not the rich) may be subject to the scripts that they should not feel resentment and should feel gratitude at handouts. Conversely, the rich may be entitled to resent a handout and not feel gratitude at a handout. These scripts could have come with conventions where the poor who express resentment at handouts are criticized and expectations of conduct like 'those in need should be subservient to their patrons'. The point in recognizing the multiplicity of scripts in (various sub-communities of) a community is to show that emotions can not only be tuned differently *across* communities, they can also be tuned differently *within a community*.

Summary. With the preceding paragraph, I can now make one final tweak to the key theses of this section: (a) emotions can underlie conventions and the expectations that come with the conventions, and (b) conventions *across and/or within communities* can tune emotions via expectations of feeling (emotional scripts) that come with the conventions.⁵⁸

⁵⁸ To avoid doubt, the examples provided thus far are non-exhaustive. I provided examples showing how those under the sway of convention A may regard convention B as wrong. For example, those who criticize others who express disgust at homosexuals may reject those who perpetuate a convention of shunning homosexuals. And those who shun homosexuals may reject those who perpetuate the convention of criticizing others who express disgust at homosexuals. This being said, there are situations where this is not the case. The EAM model does not

Tuning Moral Judgments

Now as emotions can be tuned, so also moral judgments. We saw in chapters two and three that, for EAMs, moral judgments can at least be understood as (a) psychological states that are constituted (at least in part) by emotion (or an aspect of the emotion), (b) psychological states that are based on but not constituted by emotion, (c) utterances that express emotion, (d) utterances that express a psychological state involving emotion, and (e) propositions that have emotion as referents. As emotion is essentially involved in the various types of moral judgment, the tuning of emotions by conventions can in one way or another result in the tuning of moral judgments. In this section, I intend to elaborate on this nexus between the tuning of emotions and moral judgments.

I begin with answering the question of how having emotions tuned by emotional scripts that come with conventions can result in the tuning of moral judgments *as psychological states constituted (at least in part) by emotion*. As argued in chapter two, the emotional state of seeing x as having a thick affective property plus assent, for Zagzebski, can be a moral judgment. To see x as pitiable/offensive via compassion/anger can lead to the judgment of x as pitiable/offensive. What can follow from this is a kind of *immediate* tuning of moral judgment. If an agent's compassion/anger is tuned to see y as coming under the thick affective concept of the pitiable/offensive, the agent would also be disposed to judge y as pitiable/offensive. The tuning of the agent's compassion/anger can as such result in the presence of the judgment that y is pitiable/offensive. Conversely, if an agent's compassion/anger is tuned *not* to see y as coming under the pitiable/offensive, the agent would not be disposed to judge y as pitiable/offensive. This sort of tuning results in the absence of the judgment that y is pitiable/offensive. Having emotions tuned can also result in the absence/presence of Mencian baseline judgments. As argued in chapter three, a baseline judgment can involve (for example) the perception via compassion of x as suffering, x as innocent, x as undeserving of suffering, and x 's suffering as a reason for helping x .

preclude conventions that operate in their own communal spheres without regarding another convention as wrong. A Catholic community may have the convention of intensifying alms-giving practices during Lent. Whereas a Muslim community may have the convention of intensifying such practices during Ramadan. These conventions could have arisen out of compassion. In their respective communal spheres, such intensification of alms-giving practices may even be regarded as obligatory by those who perpetuate it. However, those who perpetuate these conventions need not regard the other convention as wrong.

Unlike Zagzebski's account, the Mencian account sees such an emotion (compassion) as a judgment in its own right. However, what can still follow from both accounts is a kind of immediate tuning of baseline judgments. If an agent's compassion is tuned to see x as an object of the emotion, she can also be disposed to judge x 's suffering as a reason to help. Conversely, if an agent's compassion is tuned to *not* see x as an object of the emotion, she would not be disposed to judge x 's suffering as a reason to help.⁵⁹

With the above, we can also see that having emotions tuned by emotional scripts that come with conventions can also result in the absence/presence of moral judgments as *psychological states based on but not constituted by emotion*. An example of such a judgment is Zagzebski's account of seeing *that* x is pitiable/offensive (plus assent). As argued in chapter two, such a judgment is based on emotion in that without the prior seeing of x as pitiable/offensive, the agent is unable to see *that* x is pitiable/offensive. And seeing that x is pitiable/offensive is not a judgment that is constituted by emotion in that the agent would not have to be in the emotional state of compassion/anger to see *that* x is pitiable/offensive. Now as 'seeing that x is pitiable/offensive' is based on 'seeing x as pitiable/offensive', and the latter can be tuned by (expectations of feelings that come with) conventions, this can result in the presence/absence of moral judgments like the former. If an agent is tuned (for example) not to see x as pitiable, the agent would not be able to genuinely see that x is pitiable. We can see another example of the above judgment type in Prinz's account. As a refresher, an agent's judgment that x is wrong can for Prinz be the recognition that she is disposed to feel 'other blame' and 'self-blame' emotions when contemplating x or when x occurs. This judgment is based on emotion in the sense that it involves the agent's recognition that she has a disposition to feel anger/guilt at x when x occurs. And the judgment is not constituted by an emotion in the sense that the agent need not be in that state of anger/guilt when making such a judgment. And as an agent's judgment state of recognizing that she has a disposition to feel

⁵⁹ It is important to note that I intend to be ecumenical in this section. This means that the key points made in this section are intended to be applicable to more than one type of EAM. This also means that the use of one account of moral judgment as an example over another is not intended to show my support of one against the other. Nor is it intended that the key points only apply to the stated examples. I limit my discussion in this section to certain examples mainly for pragmatic reasons. It would be impracticable and likely to be overtly repetitive for me to consider how having emotions tuned by conventions can result in the presence/absence of every single type of moral judgment discussed so far. One example of the applicability of a key point in this paragraph to other EAMs is as follows: For Prinz, the state of anger at x (for example) can constitute the judgment that x is wrong. If an agent's disposition to anger is tuned to be elicited when she encounters x , the agent can then be disposed to judge that x is wrong. Conversely, if an agent's disposition to anger is tuned *not* to be elicited when she encounters x , the agent may not then be disposed to judge that x is wrong.

anger/guilt at x when x occurs depends on whether the agent actually has such a disposition, the tuning of the agent's emotional dispositions can result in the presence/absence of such judgment states. If x is added as an eliciting condition of the agent's anger/guilt, the agent would be disposed to feel angry/guilty at x when x occurs. This, in turn, allows the agent to make the judgment recognising that she would be disposed to anger when x occurs. Conversely, if x is removed as an eliciting condition, the agent would not be disposed to feel angry/guilty when x occurs. Consequently, the agent would *not* likely be able to judge/recognize that she would be disposed to anger when x occurs.

Moving on, having emotions tuned by conventions can also result in the absence/presence of moral judgments *as utterances that express emotion*. As an example, such judgments for Blackburn are avowals of pro-/con-attitudes. And as argued in chapter two, to genuinely judge/utter that ' x is bad' (for example) can not only be an expression of a con-attitude toward x , but can also be an expression of a pro-attitude toward those who have a con-attitude toward x and a con-attitude toward those who have a pro-attitude toward x . In keeping the example simple, to genuinely judge/utter that ' x is bad' can be an expression of anger toward x and/or those who express a pro-attitude toward x . As we can see, the judgment/utterance involved expresses the emotion of anger. And if x as an eliciting condition of an agent's anger is removed in the manner described in the preceding section, what can follow is the agent may no longer genuinely judge/utter that ' x is bad'. In this sense, we can say that having emotions tuned can result in the *absence* of such a judgment/utterance. Conversely, if y is added as an eliciting condition of an agent's anger via expectations of feelings that come with conventions, an agent would then be able to genuinely judge/utter that ' y is bad'. In this light, we can say that having emotions tuned can result in the *presence* of such a judgment/utterance. Similarly, having emotions tuned by expectations of feelings that come with conventions can also result in the absence/presence of moral judgments *as utterances that express a psychological state involving emotion*. One example of such a judgment is Gibbard's stance (as argued in chapter two) that a moral judgment can be an utterance that expresses the state of norm acceptance. We have seen that for Gibbard to judge/utter that ' x is bad' can also be understood as *it is apt to feel guilt or resentment at x* . Such a judgment/utterance expresses a noncognitive state of accepting the norms governing the elicitation of guilt and resentment. What is clear here is that the judgment/utterance does not express emotion per se. Rather, the judgment/utterance expresses *a state that involves emotion*. Without going into details of the state

involved, there appears to be room for expectations of feelings that come with conventions to govern the elicitation of resentment and guilt. Emotional scripts can add y as an apt object of resentment/guilt. If this occurs, it can result in the presence of the judgment/utterance that ' y is bad' which expresses the state of accepting the norm that y is an eliciting condition for resentment/guilt. Conversely, emotional scripts can remove z as an apt object of resentment/guilt. If this occurs, it can result in the absence of a genuine judgment/utterance that ' z is bad'.

With this section, the reader can now see more clearly as to how an EAM model can explain moral diversity: (a) emotions can underlie conventions and the expectations that come with the conventions, (b) conventions across and/or within communities can tune emotions via expectations of feeling (emotional scripts) that come with the conventions, and (c) *the tuning of emotions can result in the presence/absence of moral judgments.*

Tuning the Emotions and Moral Diversity

In the preceding section, we saw that just as emotions can be tuned, various types of moral judgments can also be tuned. As the tuning of emotions by expectations of feelings that come with conventions can result in the presence of certain moral judgments, their presence can sustain conventions or create new ones. And as the tuning of emotions by expectations of feelings that come with conventions can also result in the absence of certain moral judgments, their absence can result in the lapsing of conventions. In this section, I intend to elaborate on these points with the aid of the Mencian water and sprout metaphors outlined in chapter three.

I start with an example of a convention where some members of a community refrain from eating factory farmed animals. As argued earlier in this chapter, such a convention could have arisen out of the compassion that some agents feel for animals subject to systematic exploitation for the sake of profit. Just as King Xuan in 1A7 felt compassion for the ox led to ritual slaughter and spared it, these agents feel compassion for factory farmed animals and want to do something about their suffering. The expectations of feelings that come with this convention can in turn 'script in' factory farmed animals as an eliciting condition of compassion for other agents. And as argued in the preceding section, how compassion is tuned by expectations of feelings can result in the presence

of a certain moral judgment. In this case, the tuning involved can result in the presence of the judgment of seeing factory farmed animals as pitiable (Zagzebski's account) or seeing the suffering of factory farmed animals as a reason to act in helping ways (the Mencian account).

Such tuned judgments can be coupled with expectations of behaviour that come with the convention. To see factory farmed animals as pitiable or to see the suffering of factory farmed animals as a reason to act in helping ways can be coupled with a general expectation that factory farmed animals should not be maltreated. Such a judgment and an expectation of behaviour can offer other agents within a community normative guidance. Just as Confucian rites as 'trellises' can 'guide' an agent's sprouts (namely, an agent's disposition to act in certain ways), such a judgment and an expectation of behaviour can, as trellises, guide another agent's disposition to act in certain ways. The normative guidance involved can also be more specific. An agent can not only see the suffering of factory farmed animals as a reason to help, the agent's emotion can also be accompanied by an expectation that other agents should act in a *specific way*. In this example of a convention, it would be the expectation that other agents should refrain from eating factory farmed animals. The convention of refraining from eating factory farmed animals can be sustained in the sense that the agent's judgment plus expectation can motivate her to 'encourage' other agents to see things the way she does and to reproduce an existing pattern of behaviour of refraining from eating factory farmed animals. In short, an agent's tuned judgments coupled with her expectations of behaviour can encourage the proliferation of an existing convention.

The judgment (of seeing factory farmed animals as pitiable or seeing the suffering of factory farmed animals as a reason to act in helping ways) can also give rise to other conventions. As the judgment by itself does not pre-determine any particular behaviour, the seeing of factory farmed animals as pitiable or their suffering as a reason to help can still leave open the question of what help should the agent render. This 'opening' in turn allows for different patterns of behaviour to develop. And as argued in chapter three, agents can reflect on their involvement in the nurturing of their sprouts/dispositions. They can reflect on whether to water the sprouts, how much water is needed, etc. On this note, while expectations of feelings that come with the convention may have tuned the agent's compassion/judgment, she is still open to accept/reject the expectations of behaviour that come with the convention. She can still reflect on what help she should render (if

at all). Upon reflection and for whatever reason, the agent may instead see it fit to abstain from meat once a week. She can expect other agents to do so as well. In light of this, her judgment plus alternative expectation can motivate her to 'encourage' other agents to see things the way she does and to reproduce a pattern of behaviour of refraining from meat once a week. In other words, an agent's tuned judgments coupled with her alternative expectations of behaviour can encourage the creation of a new convention of 'meat free Mondays'.

I now continue with the example of a convention where agents who express disgust at homosexuals would be disciplined. Again and as argued earlier, such a convention could have arisen out of compassion. These expectations of feelings that come with the convention can in turn 'script out' homosexuals as an eliciting condition of the disgust of agents. Let us consider this in light of the Mencian water metaphor. An agent's disposition to see homosexuals as disgusting is akin to water flowing in a certain direction. This disposition can, however, be 'dammed up' by the expectations of feelings that come with the convention. Agents subject to the disciplinary process for example are likely to face the emotional script 'you should not feel disgust toward homosexuals'. At the disciplinary meeting, the agents may come to the realisation that it is inappropriate to see homosexuals as objects of disgust. This process can facilitate the gradual removal of homosexuals as an eliciting condition of an agent's disgust. And this is akin to the damming up of water that is flowing in a certain direction. This removal of an eliciting condition in turn can result in the absence of the judgment/seeing of homosexuals as disgusting.

The absence of such a judgment by agents can also result in the lapsing of conventions. Agents who previously saw homosexuals as disgusting may have been 'reproducing' homophobic patterns of behaviour. For example, they may have been reproducing the convention of 'condemning anyone who professes to be homosexual'. Their past judgment of seeing homosexuals as disgusting could also have been coupled with the expectation placed on other agents to reproduce homophobic patterns of behaviour. But if such agents cease to see homosexuals as disgusting, they are likely to forgo expectations of homophobic behaviour. And if more and more agents forgo such a judgment and an expectation, the convention of 'condemning anyone who professes to be homosexual' would gradually be deprived of the judgments/expectations that sustain it. As argued earlier in this section, existing expectations of

behaviour can serve as trellises that guide an agent's sprouts/dispositions. As agents who once saw homosexuals as disgusting cease to have such expectations, this cessation can also mean that there are no longer any such expectations that can serve as trellises to guide other agents' sprouts/dispositions. The convention, as a result, may slowly lapse. In this sense, conventions can lapse as a result of the absence of certain judgments.

At this stage, we can summarise the EAM model's explanation of moral diversity as follows: (a) emotions can underlie conventions and the expectations that come with the conventions, (b) conventions across and/or within communities can tune emotions via expectations of feelings (emotional scripts) that come with the conventions, (c) the tuning of emotions can result in the presence/absence of moral judgments, (d) the presence of moral judgments coupled with expectations of behaviour that come with conventions can sustain existing conventions or create new ones, and (e) the absence of moral judgments can result in the lapsing of conventions.

But there is one additional point to add to the above model before the close of this chapter, i.e. environmental factors play a role in this interaction between emotion, judgment and convention. As argued in chapter three, the Mencian sprout metaphor shows how environmental factors can influence our sprouts (emotional dispositions). Sprouts die without adequate water, sun, or the right soil. We saw that the once beautiful trees of Ox Mountain can be cut down and grazed upon in a manner that is not within the agents' control. With these in mind, we can start to see that environmental factors can influence emotional dispositions.

Resource constraints that communities are subject to can influence the eliciting conditions of an agent's emotion or whether an agent sees *x* as possessing a thick affective property. For example, when an agent herself is in need, she does not readily enter into the feelings of others, to feel their suffering, joy or resentment. This is what Adam Smith means when he argues that "[b]efore we can feel much for others, we must in some measure be at ease ourselves" (A. Smith 1790, V.I.19). Applying this point to the episode in 1A7 of the *Mencius*, if king Xuan himself was suffering from starvation due to a famine, he may not be readily able to enter into the suffering feelings of the ox. Even as the ox cries and trembles before him, the king may not be able to see (via compassion) the

ox as suffering, innocent, and its suffering as a reason to act in helping ways. He may not have said, “Spare it. I cannot bear its trembling, like one who though blameless, is being led to the execution ground” (Mencius c 300 BC: 2009, 1A7). And if we understand our sprouts as ‘first drafts’ in the sense that they are open to revision, we can see how, as a result of environmental factors, initial objects of an agent’s compassion can be removed or how an agent can no longer see certain objects as possessing the thick affective property of the pitiable. Continuing with the example provided in chapter three, the Aztecs, due to what Harner argues to be “conditions of environmental circumscription” (Harner 1977, 132), may not have been readily able to enter into the suffering feelings of their enemies. The Aztecs may as a result be unable to see the cries and trembling of their enemies as pitiable.

With the preceding paragraph, we can now modify the EAM model on the interaction between emotion, judgment and convention to the following: (a) emotions (*influenced by environmental factors*) can underlie conventions and the expectations that come with the conventions, (b) conventions across and/or within communities can tune emotions (*influenced by environmental factors*) via expectations of feelings (emotional scripts) that come with the conventions, (c) the tuning of emotions (*influenced by environmental factors*) can result in the presence/absence of moral judgments, (d) the presence of moral judgments coupled with expectations of behaviour that come with conventions can sustain existing conventions or create new ones, and (e) the absence of moral judgments can result in the lapsing of conventions.

As we can see, this EAM model can explain the diversity of moral judgments. Different emotion-based conventions can tune emotions and (simultaneously and/or ultimately) emotion-based judgments. This model can explain why agents of community A have/make certain moral judgments while agents of community B do not have/make these judgments. The model also allows us to see that (i) the presence of certain judgments within communities can *sustain and/or lapse existing conventions* within those communities, (ii) the presence of certain judgments within communities can *generate new conventions* within those communities, and (iii) the gradual absence of certain judgments within communities can result in the *lapsing of existing conventions*. In other words, the model reveals how the presence/absence of judgments can result in the presence/absence of conventions. It can explain why agents of community A have certain moral

conventions while agents of community B do not have them. With the above, I submit that the EAM model can explain moral diversity (another salient feature of morality).

Concluding Remarks

With this chapter, the reader can see that the EAM model can plausibly explain two salient features of morality (unanimity and diversity). And this explanatory power is at least one reason that counts in favour of EAMs as plausible accounts of moral judgment. As such, the novelty of my thesis here in this chapter is twofold. First, EAMs have not been used (in the manner set out above) to explain (instances of) moral unanimity and diversity. Put very briefly, we can explain instances of moral unanimity by clustering them around certain emotions. And we can explain instances of moral diversity by drawing attention to how differing emotion-based conventions can tune emotions/judgments. Secondly, the EAM model's ability to plausibly explain such salient features has never been proffered as a reason in favour of emotion-based accounts of moral judgment. In light of the above, it would be apt now for me to address an objection to the EAM model presented in the penultimate chapter of my thesis. I would also provide an additional reason in favour of EAMs in that chapter.

Chapter Five

“Gongduzi asked, ‘All are equally persons, and yet some are great persons and others are small persons – why is this?’

Mencius said, ‘Those who follow the part of themselves that is great become great persons, while those who follow the part that is small become small persons.’

Gongduzi said, ‘Since all are equally persons, why is it that some follow the part of themselves that is great while others follow the part that is small?’

Mencius said, ‘... The faculty of the mind is to think. By thinking, it apprehends; by not thinking, it fails to apprehend. This is what Heaven has given to us. If we first establish the greater part of ourselves, then the smaller part is unable to steal it away.

It is simply this that makes the great person’” (Mencius c 300 BC: 2009, 6A15).

The Exit Plan

In the previous chapter, we saw how EAMs can explain moral unanimity and diversity. On moral unanimity, having shared emotional states with shared eliciting conditions can explain why we share some moral judgments and conventions across and within cultures. On moral diversity, the ‘plasticity’ (of eliciting conditions) of emotional states can explain why some communities have agents who make certain moral judgments and other communities do not, and why some communities may have some conventions and others do not. I ended the previous chapter with the claim that the EAM model’s explanatory power is one reason that counts in favour of EAMs as plausible accounts of moral judgment.

In this penultimate chapter, I intend to address one objection to the EAM model presented, i.e. why should we accept EAMs in the first place if something non-emotional would serve as the ultimate arbiter of our moral qualms/disputes. I also intend to provide an additional reason in favour of EAMs, i.e. for EAMs, there can be more than one possible answer to a moral question. It is in this possible answers thesis that lie the novelty of this chapter. More precisely, the novelty lies in my use of Nussbaum’s right answers thesis as a starting point to the EAM model’s possible

answers thesis and how I distinguish the latter from the former. I will return to this point later, but let us start first with the objection to the EAM model.

Adjudication Problem

As argued in chapter four, the model's narrative starts with evolved emotions. Humans share some emotions in virtue of their common evolutionary past. This array of shared emotions can (in one way or another) ground judgments. These shared emotions/judgments can reveal how some conventions are shared across communities. And once we arrive at the conventional stage, we also saw how expectations of behaviour and feelings (that come with conventions) can tune emotions/judgments. This tuning process reveals how some emotions/judgments are present/absent within and across communities. One obvious implication of this is that moral qualms of an agent or disputes within and across communities are likely to happen. And one question that can be asked now is how do conscientious agents aim to resolve such disputes. The answer appears to be *by ultimately appealing to something non-emotional*. The agent can start by considering the emotions/judgments (that gave rise to the qualm or dispute) and regard them as tentative. What I mean by 'tentative' is that, upon reflection, she may reject all the emotions/judgments under consideration or accept one/some and reject the other(s). This reflective process (as we shall see later on in this section) appears to involve an appeal to something non-emotional.

It is worth noting that this appeal to something non-emotional to endorse or reject an emotion/a judgment by itself does not pose a threat to EAMs per se. After all, EAM theorists *do not* claim that only emotion is involved in resolving such disputes. As I have presupposed that some emotions can be influenced by thought (see chapter one), this opens a pathway to allow something non-emotional to influence our emotions/judgments. I shall return to this point in more detail in the next section of this chapter. At this stage, the following example would suffice. For an EAM theorist, I may see x as pitiable but reject this emotion/judgment. I may reject x as pitiable upon discovering that x was only pretending to be suffering to earn an easy buck. As such, the appeal to something non-emotional in itself is not a threat. The problem for EAMs arises when we consider the conscientious agent's apparent appeal to something non-emotional as the *ultimate*

arbiter of her moral qualms/disputes. This gives the impression that something non-emotional is doing the 'heavy-lifting' in resolving moral qualms/disputes. While the EAM model's explanatory power on moral unanimity and diversity is argued to be a reason that counts in favour of EAMs as plausible accounts of judgment, this appeal to something non-emotional as ultimate arbiter appears to be a reason that counts against EAMs. Such a position raises a fundamental issue of why should we accept EAMs in the first place if something non-emotional would serve as the ultimate arbiter of our moral qualms/disputes. This is the objection that I aim to address.

Cultural tunings. Allow me to start by considering how there can be conflicting emotions/judgments between agents across and within communities. Elster argues that "[w]hereas all animals of a given species have pretty much the same emotions, human emotions vary heavily with local culture... culture affects emotional life in a more substantive way by defining the *behaviors* that are appropriately viewed as objects of contempt and shame. Many spontaneous actions that would pass unnoticed in some societies can trigger strong emotions in others" (Elster 1999, 411-412). The overarching theme here is that culture exercises a heavy influence on our emotions. While Elster's focus is on *contempt and shame*, we can extend his point to *emotions in general*. And while Elster's emphasis is on the *behaviour of agents* as objects of emotions, we can also extend his point to other objects like *agents themselves*. I can (for example) see the behaviour of an agent as disgusting or the agent herself as disgusting. In a similar vein, Charles Taylor argues that:

"... it only makes sense to see these [my shrill voice, or my effeminate hands] as objects of shame if they have for me or my culture an expressive dimension: a shrill voice is (to me, to my culture) something unmanly, betokens hysteria, not something solid, strong, macho, self-contained. It does not radiate a sense of strength, capacity, superiority. Effeminate hands are – effeminate. Both voice and hands clash with what I aspire to be, feel that my dignity demands that I be, as a person, a presence among others" (Taylor 1985, 53).

For Taylor, culture plays a role in determining appropriate objects of emotion. If shrill voices are not culturally regarded as unmanly or if effeminate hands not regarded as effeminate, the agent may not see having a shrill voice or effeminate hands as appropriate objects of shame. In sum, it appears that what unites Elster and Taylor in the above two passages is that culture determines what counts as appropriate objects of emotions.

Now let us apply the above point to what I argued in the previous chapter. As argued, an emotional script can come with a convention. For example, the convention of barring Jews from public schools in pre-war Nazi Germany can come with a script 'you should be disgusted by the Jews'. This script in turn can tune the disgust of agents within Nazi Germany. Agents (who did not previously see Jews as disgusting) may come to see Jews as disgusting under the influence of this script. And if we accept that scripts are "norms that govern emotion... drawn from a particular, historically shaped, linguistic and cultural repertoire" (Pettigrove and Parsons 2012, 509), we can see how culture can define what counts as appropriate objects of emotions *via emotional scripts*. It would be scripts (drawn from a network of historically shaped cultural values and beliefs) that tune emotions. The script 'you should be disgusted by Jews' can (for example) be drawn from the cultural belief of some agents within Nazi Germany that the Jews are "insidiously subhuman" and "[t]heir ostensible humanity was, at best, only skin deep" (D. L. Smith 2011, 5). Allow me to be even more specific. Such a script may have been drawn from cultural beliefs articulated in *The Subhuman* (edited by Himmler and circulated in Nazi Germany in 1942) where it was argued that the Jews were "beasts in human form" and:

"Just as the night rises against the day, the light and dark are in eternal conflict. So too, is the subhuman the greatest enemy of the dominant species on earth, mankind. The subhuman is a biological creature, crafted by nature, which has hands, legs, eyes, and mouth, even the semblance of a brain. Nevertheless, this terrible creature is only a partial human being.... Not all of those who appear human are in fact so" (D. L. Smith 2011, 155).

With the above, we see that culture can tune emotions via emotional scripts.⁶⁰ And as we differ in cultural scripts within and across communities, what can follow is that agents within and across communities can have differing and even conflicting emotions/judgments. While Nazi agents see Jews as disgusting, other agents do not. Instead, other agents see Nazi agents (who see Jews as disgusting) as offensive. In turn, Nazi agents may see other agents (who see Nazi agents as offensive) as offensive. The qualm that a Nazi agent may face or the dispute between them is whether Jews should be seen as disgusting. Here is a more contemporary and relatable example. While pro-lifers judge that 'abortion is wrong' (an expression, perhaps, of their anger toward acts

⁶⁰ We saw in chapter four how such expectations of feeling can be emotion-based. Such scripts can be sustained by emotions. In a similar vein, a cultural belief that the Jews are beasts in human form can also be sustained by emotions such as anger, disgust, etc.

of abortion), pro-choicers do not. Instead, pro-choicers may judge that 'restricting abortion is wrong' (an expression of anger toward such restrictions). And the qualm that pro-lifers/choicers may face or the dispute here is whether abortion is wrong. The questions now are how do conscientious agents (a) who are parties to such disputes justify their respective positions, and (b) who are non-parties to such disputes adjudicate them.

Disputants ultimately appealing to something non-emotional. Peter Singer argues that "[u]nlike the other social mammals, we can reflect on our emotional response, and choose to reject them... Undoubtedly, [Kant] was mistaken to think that morality can be based on reason alone, but it is equally mistaken to see morality only as a matter of emotional or instinctive responses, unchecked by our capacity for critical reasoning" (Singer 2006, 149-150). In a similar vein, Christine Korsgaard argues that:

"... as rational beings we are aware of our attitudes. We know of ourselves that we ... fear certain things, love certain things ... and so on... We are aware of the *potential grounds* of our actions – of the ways in which our attitudes incline us to respond. And once you are aware of the influence of a potential ground of action, you are in a position to decide whether to allow yourself to be influenced in that way or not... you now have a reflective distance from the impulse that is influencing you, and you are in a position to ask yourself, 'but *should* I be influenced in that way?' You are now in a position to raise a normative question, a question about whether the action you find yourself inclined to perform is justified" (Korsgaard 2010, 18).

With the above in mind, we see that Singer and Korsgaard are united in the claim that our emotions/judgments are tentative in the sense that we can reflect on them and subsequently reject them. The implicit claim is also that while our evolved emotions/judgments may have given rise to morality, these emotions/judgments do not have a final say. One of the trolley problems can help me elucidate this point. Imagine that:

"you're on a footbridge overlooking the railway track. You see the trolley hurtling along the track and, ahead of it, five people tied to the rails... There's a very fat man leaning over the railing watching the trolley. If you were to push him over the footbridge he would tumble down and smash on the track below. He's so obese that his bulk would bring the

trolley to a juddering halt. Sadly, the process would kill the fat man. But it would save the other five... Should you kill the fat man?" (Edmonds 2014, 36)

A consequentialist may answer this question by saying that while she has a con-attitude against pushing the fat man off the footbridge to save five, this con-attitude/judgment is an unreliable guide to the question what should I do. Instead, she may consider that it is better that one die for the lives of the many. And while a deontologist may also say that she has a con-attitude against pushing the fat man off, she may upon reflection endorse this con-attitude. She might say that we should never treat anyone as a means to an end and her emotional aversion hints toward or reveals this principle. Here we see that the non-emotional principle is prioritized over the emotion. The point I would like to emphasize here is that at least in this instance the consequentialist and deontologist acknowledge the influence of their emotions/judgments on the question 'what should I do' but appeal to something non-emotional to *ultimately* justify their respective positions.⁶¹

With a hypothetical example out of the way, let us consider now a more realistic example. Agents may see homosexual acts as disgusting/wrong. While they may acknowledge the emotion's influence on what is right/wrong, they may diverge on the way they make sense of their disgust. Post-reflection on her seeing of homosexual acts as disgust/wrong, agent *A* may upon reflection endorse her disgust as a reliable guide to the question of what counts as wrong. Disgust for her "marks out moral matters for which we can have no compromise" (Miller 1997, 194). She may consider disgust to be "the emotional expression of deep wisdom... warning us not to transgress what is unspeakably profound" and that the emotion may be "the only voice left that speaks up to defend the central core of our humanity" (Kass 1998, 687). In the case of her disgust/judgment toward homosexual acts, she may consider this as a warning not to transgress the 'natural purpose' of sexual relationships. To the contrary, Agent *B* may consider her seeing of homosexual acts as disgusting/wrong to be an *unreliable* guide to the question what counts as wrong. She may consider her disgust misguided as "its thought-content is typically unreasonable, embodying magical ideas of contamination, and impossible aspirations to purity, immortality, and nonanimality, that are

⁶¹ Though I use the term disputants implying that the consequentialist and deontologist encountered each other and are disputing about whether it is alright to kill one to save five, there is no need for such a dispute to prompt an agent to reflect on her emotion/judgment. A consequentialist may have 'encountered' the deontologist's position when she reflects on her own judgment and positions that run contrary to her judgment. In such an instance, she is not engaging in a dispute. Rather, she may be experiencing a moral qualm, i.e. a doubt as to whether her initial emotion/judgment is the right one.

not just in line with human life as we know it" (Nussbaum 2004, 14). In other words, she may reject her disgust/judgment on grounds that it involves 'magical thinking' and any such emotion/judgment should not be rejected. With the above, it appears that while agents *A* and *B* ultimately diverge on whether homosexual acts are wrong, they are united in their appeal to something non-emotional to ultimately justify their respective positions. Agent *A* is essentially saying that her seeing of *x* as disgusting hints her toward the 'natural' purpose of sexual relationships and how this natural purpose should not be violated. It is not her disgust that makes homosexual acts wrong. Rather, it is how homosexual acts violates the natural order of things that makes them wrong. As for agent *B*, she is essentially saying that while seeing of *x* as disgusting *can* influence our moral thinking, we should keep such influence in check. And she appeals to some other non-emotional source to justify her position that homosexual acts are not wrong.

With the above, we see that agents who share the same emotion/judgment can diverge on their interpretations of it. The divergence reveals an appeal to something non-emotional as the ultimate arbiter to endorse/reject the emotion/judgment. This appeal to the non-emotional applies also to situations where agents have differing attitudes toward the same situation(s). While the deontologist may continue to have a con-attitude against pushing the fat man off the footbridge, the consequentialist may no longer have such a con-attitude. Instead, the consequentialist may have a con-attitude toward the deontologist's endorsement of her emotion/judgment. And the deontologist may have a con-attitude toward the consequentialist's rejection of her emotion/judgment. Both the consequentialist and deontologist can attempt to justify their differing con-attitudes on the same subject by appealing to something non-emotional. Similarly, while agent *A* may continue to see homosexual acts as disgusting, agent *B* may no longer see them as such. Instead, agent *A* may have a pro-attitude toward those who see such acts as disgusting. And agent *B* may have a con-attitude toward this pro-attitude. Again, agent *A* can attempt to justify her pro-attitude and agent *B* her con-attitude by appealing to something non-emotional.⁶²

⁶² It is worth noting that the examples in this and the preceding two paragraphs are only some ways which agents can reflect on their emotions/judgments. There are many ways which agents can reflect on their emotions/judgments. Furthermore, these examples also reveal only a sliver of what EAM theorists regard moral judgments. As argued in chapter two, emotions can be moral judgments in themselves for some. Moral judgments can express emotions for others, etc. It would be impractical and largely repetitive to consider how each type of moral judgment can be reflected upon and subsequently endorsed/rejected. Regardless, the point remains that it appears that agents can reflect on their emotions/judgments and ultimately appeal to something non-emotional to endorse or reject them.

Adjudicators ultimately appeal to something non-emotional. The above nicely segues us into how agents who are non-parties to such qualms/disputes can adjudicate between differing moral positions. John Rawls in outlining a decision procedure for ethics asks how should agents “determine the manner in which competing interests should be adjudicated, and, in instances of conflict, one interest given preference over another...” (Rawls 1951, 177). His project is a normative one, and it focuses on competing interests instead of competing emotions/judgments. However, some tweaks can be made to fit the context. I can reframe the question as follows: *how do non-party agents adjudicate between competing emotions/judgments, and in conflicting instances prefer one emotion/judgment over another.* At a high level, the answer appears to be that competing emotions/judgments are adjudicated (again) by ultimately appealing to something non-emotional. Just as conscientious parties to the dispute appeal to something non-emotional to endorse/reject their emotions/judgments, conscientious adjudicators appear to appeal to something non-emotional to resolve the dispute. Rawls, for example, argues that the adjudicator’s decision must be “one which has been preceded by a careful inquiry into the facts of the question at issue, and by a fair opportunity for all concerned to state their side of the case” (Rawls 1951, 182). He continues to argue that the adjudicator’s decision must also be “intuitive with respect to ethical principles... [one that is] not determined by a systematic and conscious use of ethical principles” (Rawls 1951, 183). If we grant this, what can follow is that the adjudicator would consider their pro- or con-attitudes of the consequentialist and deontologist, agents *A* and *B*, and their reasoning behind such attitudes, before deciding on them. The adjudicator may, upon reflecting on their attitudes and reasoning behind them, accept one and reject the other. She may consider (for example) that the consequentialist position has extreme implications and thus reject this position. Crudely put, she may have implicitly applied, inter alia, the principle ‘a moral position should not have extreme implications’ in coming to her decision. The point here is that the adjudicator appears to ultimately appeal to something non-emotional to adjudicate between differing moral positions.

Of course, an adjudicator need not adopt Rawls’ decision making procedure to resolve disputes. She may opt to explicitly adopt a substantive (as opposed to procedural) ethical framework to resolve disputes. She may be a Kantian and adopt Korsgaard’s version of the universalizability

test, namely a maxim should be rejected if we are unable to will it as law that “all rational beings could agree to act on together” (Korsgaard 1996: 2010, 99). And in adjudicating the dispute between the consequentialist and Kantian, she may reject the former’s position on grounds that the maxim ‘you shall always kill one to save five’ cannot be willed as a universal law. Such a maxim is not something that all rational beings can agree to act on together. In addition, the adjudicator may even apply a religious code to resolve dispute. Like a judge who is bound to interpret and apply the law to disputes, the adjudicator who is an adherent to a ‘utilitarian god’ commanding the application of utilitarian principles may reject the Kantian position in favour of the consequentialist position instead. John Austin may be one such adjudicator as he argues that “[t]he commands which God has revealed we must gather from the terms wherein they are promulg[ate]d. The command which he has not revealed, we must construe by the principle of utility: by the probable effects of our conduct on that general happiness or good which is the final cause or purpose of the good and wise lawgiver in all his laws and commandments” (Austin 1995: 2001, 94). The point again is that the adjudicator appears to ultimately appeal to something non-emotional (a decision-making process, substantive ethical framework, religious code, etc.) to resolve disputes.

With the above, we see agents (who may be parties to a dispute) and adjudicators (who aim to resolve such a dispute) appear to ultimately appeal to something non-emotional to resolve a dispute. This appears to count as a reason against EAMs as plausible accounts of moral judgments. Moral judgments as outlined in chapter two appear only to be *pro-tanto judgments*. They remain subject to revision by a ‘non-emotional’ reflective process where an agent would draw on non-emotional principles, procedural ethical frameworks, substantive ethical frameworks, or religious frameworks. This raises the fundamental issue of why consider EAMs in the first place if emotion-based judgments are ultimately subject to something non-emotional. If left unaddressed, EAMs may lose out in this respect. As a shorthand, I call this the adjudication problem.

Reply to the Adjudication Problem

My reply to the adjudication problem comprises four steps. The first step involves accepting that emotions/judgments can be rejected/endorsed upon reflection. The second step then involves

accepting that the agent can draw on *non-emotional sources* within this reflective process. The third step is where the action largely lies. This step involves the argument that while the agent can (within the reflective process) draw on something non-emotional, various reflective judgments can remain emotion-based in at least the following ways:

- The agent's emotion/judgment (i.e. the seeing of x as having a thick affective property, emoting at x) itself could have resulted from her reflection.
- The agent's reflective judgment that x has a thick affective property/ x should be seen as having a thick affective property can be derived from her reflective judgment of seeing x as having a thick affective property *or* can be based on the (implicit) recognition that she is disposed to feel an emotion toward x .
- The agent's reflective judgment that x does not have a thick affective property/ x should not be seen as having a thick affective property refers to a corresponding emotion *or* refers to the agent's (disposition to) an emotion *or* can 'point to' other emotion-based reflective judgments.

The last and fourth step involves the claim that the agent's reflective process can also be emotion-based. The process is emotion-based in the sense that an agent can draw on other emotions (alongside non-emotional sources) within the reflective process. With these steps in mind, let us start with the first.

Reflection, endorsement, and rejection. For EAMs, an agent can have a pro-/con-attitude toward x . The agent can also reflect on this pro-/con-attitude toward x and subsequently endorse or reject the attitude. Let us consider three examples, with each emphasizing on different aspects/outcomes of reflection. I start with Zagzebski. As argued in chapter two, the seeing of x as offensive for Zagzebski is constitutive of a judgment. For her, this con-attitude toward x can also be reflected upon. She argues that:

“When I say ‘She is pitiful,’ I am willing to assert that the intentional object of my pity has the property of being pitiful, whereas if I only say ‘I pity her,’ I am calling attention to myself and my emotion, and this may indicate that I am unwilling to say that I have construed the intentional object of my emotion correctly. It is even possible to say ‘I pity her, but she is not pitiful.’ We cannot make statements like that without *higher-order*

reflectiveness on our own emotional states and their intentional objects [emphasis added]" (Zagzebski 2004, 169).

I draw three points from this argument. First, the statements an agent can make reveal her ability to distinguish between the descriptive claim that she has a con-attitude and normative claim that she should (not) have a con-attitude.⁶³ An agent may see x as offensive but reject this as an incorrect perception. Second, the agent's ability to distinguish between the descriptive and normative claims reveals her ability to reflect on her con-attitude. Third, the agent's reflection can involve considering the question 'should I see x as offensive'. In asking and answering this question, the agent can endorse or reject her con-attitude. Zagzebski argues that "[t]here are feedback mechanisms that change our emotions when the consequences of acting on them are not expected. We also revise by imitation of the emotions admired by others. Literature is helpful in this respect... Not only do we have our own emotional response to a character's situation, our awareness of being outside the story also allows us to critically compare our own emotions and the emotions of the characters in the story" (Zagzebski 2004, 92). I gather from this passage that an agent upon reflection can consider how others would react to her con-attitude and subsequently endorse/reject it on this basis. The agent can also compare her con-attitude toward x to how exemplars would react toward x and endorse/reject her con-attitude accordingly. In short, we see that for Zagzebski a con-attitude felt can be endorsed or rejected upon reflection.

Now let us turn to Blackburn. As argued in chapter two, for Blackburn, the judgment that x has a negative moral attribute can be an avowal of a practical state (emotional states included). He further argues that such practical stances are "reason-responsive": "There are cases in which we can simply cause impulses and desires to vanish by reflection on things, although there are also cases in which we cannot escape the force of desire or emotion so easily... Imaginations, beliefs, and chains of reasoning can all blunt their force" (Blackburn 2002, 123). I will focus on two points on Blackburn's position. First, an agent can reflect on her con-attitude toward x by drawing on a network of beliefs and/or reasoning. Let us consider a simple example. The agent may start off by expressing the con-attitude that 'John is bad'. She can then reflect on this attitude by asking 'am I

⁶³ Such a distinction can also be seen implicitly in Gibbardian moral judgments. As argued in chapter two, for Gibbard, to claim that x has a negative moral attribute is to claim that it is apt to feel guilt and/or resentment toward x . What is implicit in such a moral judgment is that there is a distinction between the descriptive claim that 'I feel resentment toward x ' and the normative claim that 'it is apt or inapt to feel resentment toward x '.

biased in my judgment against John as he got the job I really wanted and interviewed for'. Implicit in this question can be the underlying belief that con-attitudes should at the very least be impartial. The agent can proceed to draw on her beliefs about John to assess her con-attitude. She may upon reflection continue to endorse her con-attitude (John got the job because he lied about his CV) or reject her con-attitude (John got the job because he is better qualified and more experienced). Second, the agent's reflection can eradicate or suppress her con-attitude toward x . Upon reflection and after considering that John got the job fair and square, the agent's con-attitude toward John may 'vanish'. This, however, need not be the case. Even if the agent rejects her con-attitude toward John, this attitude may still remain. She may recognize that she should not but cannot help feeling this way. As Blackburn argues, "a consideration may continue to produce its affect, even when we wish it wouldn't or deceive ourselves into thinking that we have neutralized it" (Blackburn 1998, 132). In such a case, she may suppress her con-attitude by refusing to express it. Here we see how Blackburn's position adds on to Zagzebski's position above: not only can a con-attitude felt be endorsed or rejected upon reflection, rejecting a con-attitude upon reflection can also result in the eradication or suppression of the con-attitude. And while Blackburn's passage above is silent on whether the endorsement of a con-attitude can result in the intensification of a con-attitude, I do not see why this cannot be the case. In other words, it appears that upon reflection and after accepting that John cheated his way toward getting the job, the agent's con-attitude toward John may intensify. The key point here is that not only do EAM theorists accept that agents can reflect on their emotions/judgments, they also accept that agents' reflection can influence their emotions/judgments.

Now let us turn our attention to Slote, our third and last example. As argued in chapter two, for Slote, moral judgments as propositions (expressible via sentences and/or utterances) have empathic warmth/chill and what the empathic warmth/chill is about as referents. The term 'bad' in the judgment 'Hitler is bad' can refer to an agent's empathic chill toward Hitler's 'cold' treatment of the Jews. An agent, however, who is personally involved in the situation can fail to feel empathic chill or confuse her other phenomenologically similar emotional experiences as empathic chill: "... a mother whose son is on death row for murdering a baby may be so concerned about saving him from death that she fails to experience the feelings of disapproval she would feel about another killer... the baby's parents may be too angry as a result of their personal loss for their disapproval of the killer to register as a distinct and separate phenomenon" (Slote 2010, 41).

In other words, an agent's intense emotional experiences (when personally involved in a situation) can 'smother' or be confused as empathic chill. To manage this problem, the agent can adopt an impartial standpoint. Slote argues that his account "treats conditions of impartiality... as helping to (epistemically) *clarify* whether given warmth or coldness really constitutes approval or disapproval and ... as *facilitating* the actual occurrence of these moral reactions" (Slote 2010, 42). While Slote does not go into details of *how* this clarification/facilitation would work, I consider *reflection* to be a plausible way of distinguishing empathic chill from other emotional experiences and encouraging the feeling of empathic chill. The agent can adopt a third-person perspective and ask 'what would I feel if I were not personally involved in the situation'. She can engage her imagination in considering an answer to this question. On one hand, she might imagine the murderer as a stranger (as opposed to her son) and consider how she would feel. In this case, she may feel empathic chill toward the murderer. She may realize that her concern for her son has 'smothered' the empathic chill that she would have felt if the murderer was not her son. She may then upon reflection consider that she would feel empathic chill toward a murderer who is not her son and (*ceteris paribus*) should feel empathic chill toward the murderer son of hers. On the other hand, an agent might also imagine the victim as someone else's child (as opposed to her own). In this case, while she may still feel anger at the murderer, the intensity involved may not be as great. She may then realise that her anger/disapproval at the murderer of her child might have 'smothered' her empathic chill. Like the first instance, the agent may subsequently on reflection consider that she would feel empathic chill toward the murderer if the child was not hers and (*ceteris paribus*) should feel empathic chill toward the murderer of their child. It is in this sense that the agent can (upon reflection) endorse an empathic chill toward her murderer son/the person who murdered her child. The key point to note here is that a Slotean account highlights a different aspect of how reflection can be involved in moral judgments. Unlike Zagzebski's and Blackburn's positions, the agent does not upon reflection endorse a con-attitude that is *felt*. Rather, she endorses a con-attitude that is *absent* but *should* be felt.

Drawing on non-emotional sources. With the above, we see that at least some EAMs recognize that con-attitudes (felt/absent) can be rejected/endorsed upon reflection. Reflection can also result in the eradication, suppression and intensification of con-attitudes. The next step of my argument involves the claim that at least some EAMs are sufficiently sophisticated to incorporate the claim that agents can draw on what appears to be non-emotional sources within this reflective process.

Let us consider two examples, with each emphasizing on different 'non-emotional' sources that can be drawn upon by the agent.

I start again with Zagzebski. She argues that "[a] good emotion, like a true belief, fits what it is about" and a good emotion can be defined "as the emotion an exemplar would have in a given situation" (Zagzebski 2004, 78, 83). And as argued in chapter two, an emotion (plus assent) for Zagzebski can also be a moral judgment. With this in mind, we can say that a good emotion/judgment is one that fits its intentional object. The seeing/judging of x as pitiable is fitting if and only if x possesses the thick affective property of the pitiable. If there is a dispute on whether x is pitiable, agents can be guided by whether exemplars would see/judge x as pitiable. At this point, one may consider that even if agents agree on the exemplars, exemplars can still differ on whether they see x as pitiable. How then do we settle the dispute? To answer this question, Zagzebski is likely to point to God as the "supreme exemplar" (Zagzebski 2004, 226). She argues that "[a]ll moral properties are ultimately grounded in the emotional experience of God. The value of a human emotion in a given situation is limited by the emotions that God would have if He were in that situation" (Zagzebski 2004, 226). In other words and crudely put, if God was in our situation and sees x as pitiable, then x is pitiable.⁶⁴ On this note, we can start to see how an agent can (upon reflection) appeal to a religious framework to endorse/reject her con-attitude. An agent may see x as offensive. In being aware that this emotion might be wrong, the agent may then reflect on how God the supreme exemplar would feel in her situation. If she considers that God would (not) see x as offensive, she would endorse/reject her emotion. The principle that the agent appears to be applying can be roughly framed as *when in doubt consider how God would feel/judge in my stead*.

Now let us turn to Prinz. He argues that he does not want "to insist that moral attitudes typically arise in the absence of deliberative reasoning [as]... it may take a lot of inference before we see an action in a way that triggers an emotional response... In some cases, emotions require a fair amount of deliberation before they arise. We may have to think about a foreign policy decision for hours before it elicits condemnation or praise" (Prinz 2007, 98, 114). And as argued in chapter two, an agent can for Prinz deliberate on whether x comes under category y (a category of objects where

⁶⁴ To avoid doubt, I have not and will not be assessing the merits of Zagzebski's position. The point of raising this is to consider how an agent can upon reflection appeal to a religious framework to endorse/reject her con-attitude.

she would be disposed to feel anger, disgust or contempt at). If the agent deems that x comes under y , her anger/disgust/contempt at x can be triggered.

The first point to note from the outset is that reflection for Prinz can occur prior to and result in the judgment that x is wrong. The second point is that Prinz's model appears to allow the agent to draw on a whole range of what appears to be non-emotional sources when she reflects on whether x comes under y or is wrong. For example, an agent may carefully reflect on whether abortion (x) comes under the category of murder (y). If she is disposed to feel anger at murder and if she (upon reflection) determines that abortion is murder, she would also be disposed to feel anger at abortion/judge that abortion is wrong. The agent can consider the question whether abortion is wrong from various vantage points. Is an embryo/zygote a person? Does a human embryo/zygote have the right to life? Should the pregnant woman have a free rein over her body? These are some issues that an agent might consider. And they may be based on beliefs on what can or cannot be done to a person, what rights are unqualified/qualified, etc.

The agent may be a preference utilitarian. If she considers that "an action contrary to the preference of any being is, unless this preference is outweighed by contrary preferences, wrong", she may also consider that "taking the life of a person will normally be worse than taking the life of some other being, since persons are highly future-oriented in their preferences" (Singer 1993, 94-95). She may as a result consider that killing a person is wrong. She can as such be disposed to feel anger at those who murder/judge that murder is wrong. In the abortion debate, the agent may have to consider whether an embryo/zygote is a person. On this point, the agent may consider that an embryo/zygote is not a person as it is not self-conscious, it does not have preferences let alone highly future-oriented ones. With such reflection, the agent may reject the claim that abortion is wrong and may not be disposed to feel anger at those who procure abortions. The agent may instead be disposed to feel anger at those who do not allow abortions as they deny the preferences of some women without any justifiable reason. With this in mind, we see that the agent can draw on a normative ethical framework in her deliberations that can result in anger at pro-lifers/the judgment that prohibiting abortion is wrong.

Still grounded in the emotions. Zagzebski's and Prinz's positions allow us to see that at least some EAM theorists anticipate that agents can draw on what appears to be non-emotional sources (e.g. religious beliefs, beliefs on personhood, normative ethical theories, etc.) within this reflective process. We can also generalize their positions in the following manner: agents can draw on *whatever* non-emotional sources while reflecting on whether x has a moral attribute. Agents can assess whether x has a negative moral attribute based on whatever principles that they see fit. An agent can (for example) adopt the approach that x has a negative moral attribute if and only if it does not contradict belief(s) a , b , c , d , and/or e . Furthermore, Flanagan appears to take the term *whatever* seriously and argues that "[w]hat is relevant to ethical reflection is everything we know, everything we can bring to ethical conversation that merits attention – data from the human sciences, from history, from literature and the other arts, from playing with possible worlds in imagination, and from everyday commentary on everyday events" (Flanagan 2016, 44). The point to be emphasized here is that EAM theorists can grant all these while insisting that judgments resulting from reflection are still grounded in the emotions. More specifically, while agents/adjudicators appeal to something non-emotional to resolve a qualm/dispute, their reflective judgments are still in one way or another grounded in the emotions. By making this move, the question why consider EAMs in the first place if emotion-based moral judgments are subject to something non-emotional becomes a non-issue. Reflective judgments can still be emotion-based. On top of this, the move also reveals to the reader the 'versatility' of at least some EAMs in that they can account for agents' appeal to reflection and non-emotional sources to resolve moral qualms and/or disputes.

I start with Blackburn. He argues that "our attitudes and practical stances need discussion. They put us in conflict with each other, or in conflict with ourselves. Even simple prescriptions need discussion and defense, and the overall field of obligations and values is much more complex still" (Blackburn 2002, 127). We can read this need for discussion in light of what was argued in chapter two, i.e. for Blackburn, we risk violent conflict if we do not coordinate our attitudes. Blackburn's main thrust here is to justify the propositional surface of moral discourse (despite the non-cognitive status of our moral judgments). But in our context, we can recruit his point for a different purpose. What I draw from Blackburn is that agents who reflect on their attitudes are more likely to be able to defend them or provide reasons for them. And adjudicators who reflect on their own attitudes and those of the disputants are better able to discuss and resolve moral disputes. Put this

way, reflection for Blackburn is not only part of his account of moral judgment, it is also *encouraged*. And more importantly, the agent's/adjudicator's reflective process (and her drawing on "resources of logic" to resolve moral qualms/disputes (Blackburn 2002, 127)) need not result in detracting from the claim that reflective judgments can remain grounded in the emotions. Let us now consider more specifically how such reflective judgments can still be emotion-based.

As argued above, for Zagzebski, an agent may see *x* as offensive and may then reflect on how God the supreme exemplar would feel in her situation. If the agent considers that God in her stead would see *x* as offensive, she would endorse her own seeing of *x* as offensive. The agent's seeing of *x* as offensive can thus become a reflective judgment in the sense that the emotion is one that is not just felt, but also reflected upon and endorsed by the agent. Such a reflective judgment obviously remains emotion-based. Reflective anger (as opposed to a gut reaction sort of anger) is still anger. One may say however that the agent (post-reflection and endorsement of her anger toward *x*) would not always see *x* as offensive (or feel anger toward *x*). When an agent encounters *x*, she would not always experience an emotional state of anger toward *x*. Instead, the agent may see *that x* is offensive. If so, perhaps such a judgment *that x* is offensive (post-reflection) is not emotion-based. To the contrary, a judgment of this sort can still be indirectly emotion-based. What I mean by this is seeing *that x* is offensive is derived from the seeing of *x* as offensive. And as argued in chapter two, an agent can only genuinely make the judgment *that x* is offensive if she had a prior judgment of seeing *x* as offensive. Considering this, an agent's judgment *that x* is offensive (post-reflection) can still be emotion-based.

The agent's endorsement of her seeing of *x* as offensive can also be understood or framed as the judgment that *x* *should* be seen as offensive. Krister Bykvist's point on the fitting attitude analysis of value can be helpful here. He argues that "[*x*] is good = df. [*x*] is such that it would be fitting to favour [*x*], if one were to contemplate [*x*]" (Bykvist 2009, 4). If we replace 'good' with 'offensive' and 'favour' with 'disfavour', we can see that to think that *x* is offensive is to think that it would be fitting to disfavour *x* upon reflecting on *x*. We can also see that the is-statement involved (*x* is offensive) is not understood as *x* is offensive to me (merely reporting my feelings toward *x*). Nor is the statement understood as *I would be disposed to feel anger toward x* (merely predicting my feelings toward *x*). Instead, the statement *x* is offensive is understood as an endorsement of the anger that

is/would or could be directed at x . As Daniel Jacobson argues, [t]o call something funny is in some way to endorse amusement at it, not to report or predict it" (Jacobson 2011). With this in mind, we can also say that to think that x is offensive can be a way to endorse anger at x . We can probe more into what this endorsement can be about. Justin D'Arms and Jacobson argue that "to think an emotion a fitting response to some object is to think there is (pro tanto) reason, of a distinctive sort, for feeling the emotion toward it" (D'Arms and Jacobson 2006, 108). In this spirit, the agent's judgment that x should be seen as offensive can be the result of a deliberative process where reasons are put forward in support of feeling anger toward x .⁶⁵ A question that can be asked now is whether this sort of judgment can be emotion-based. I would say 'yes'. The judgment that x should be seen as offensive involves a reference to the thick affective property of the offensive. And such a judgment can be emotion-based in the sense that it refers to a thick affective property that corresponds with the emotion of anger. But as Zagzebski argues, "[a] person can be wrong in her construal of something as ... contemptible or rude. She might see something as contemptible that is not contemptible; she might see something as offensive that is not offensive, and so on. And she can come to recognize her mistake" (Zagzebski 2004, 77). In this light, if the agent reflects on her seeing of x as offensive and comes to reject her emotion/judgment, this rejection can be framed as the judgment that x should not be seen as offensive. This judgment can also be emotion-based in the sense that its content refers to a thick affective property that has a corresponding emotion.

One may at this point object that such a reference is insufficient for the reflective judgment to be emotion-based. More precisely, it would be insufficient for the reflective judgment that x should not be seen as offensive to be emotion-based as a judgment could be about an emotion without being emotion-based. One response to this is that such a reflective judgment can still be emotion-based in the sense that the judgment that x should not be seen as offensive can be the result of seeing x as having another thick affective property or other thick affective properties. For example, an agent may initially see Jean Valjean's stealing of a loaf of bread as offensive. But she may (upon reflection) judge that Valjean's theft should not be seen as offensive. The basis of her judgment may be the result of her reflection on the wider context of Valjean's theft. She may have discovered that

⁶⁵ It is worth noting that D'Arms and Jacobson admit only a specific class of reasons for feeling. They argue that "[r]easons of fit are those reasons that speak directly to what one takes the emotion to be concerned with, as opposed to reasons that speak to the advisability or propriety of having that emotion" (D'Arms and Jacobson 2006, 108). I remain neutral on this point. But regardless of whether the class of reasons should be restricted, the judgments that result from such deliberations can still remain emotion-based.

Valjean stole a loaf of bread in an economic downturn for his sister's starving children. She may have (as a result) seen Valjean's motivation behind his theft as admirable. He stole not for his own sake but for others who are starving. She may also have seen Valjean's nieces'/nephews' starvation as pitiable. In other words, the agent's judgment that *x should not be seen as offensive* can still be emotion-based in the sense that it is based on her seeing of *x* as admirable and seeing of *y* as pitiable.

Now let us consider how the adjudicator would resolve some moral issues between disputants. Disputant A may see *x* as offensive and put forward the claims that *x* is offensive and should be seen as offensive. To the contrary, disputant B may not see *x* as offensive and reject these claims. The adjudicator would now have to resolve this dispute, and a key issue to resolve is whether *x* should be seen as offensive. In such an instance, the adjudicator may (not), on a preliminary basis and prior to considering the positions of the disputants, see *x* as offensive. This preliminary position of the adjudicator may (not) be a reflective judgment. The adjudicator may then reflect on the reasons put forward by the disputants and come to a determination. She may upon reflection determine that *x* should (not) be seen as offensive. Regardless of what the adjudicator determines, her reflective judgment can still be emotion-based. At the very least, the content of her judgment would still refer to a thick affective property. In sum, with the above we see that at least some of the agent's reflective judgments can still be emotion-based even if she draws on non-emotional sources in the reflective process. The same applies to the reflective judgments of adjudicators.

Prinz offers a different perspective on how reflective judgments can still remain emotion-based. He argues that there are "grounding norms". They are "norms that we tend to regard as not needing explanation. In our value system, examples include prohibitions against killing, stealing, and incest. When someone asks "Why is killing wrong?" we respond with an incredulous stare or fumble through a *post hoc* explanation. Only lunatics and philosophers ask such questions" (Prinz 2007, 126). This point is important to draw to the reader's attention as it reveals another aspect of how reflective judgments can remain emotion-based while accounting for an agent's reflective process (and drawing on non-emotional sources). Allow me to elaborate. Suppose an agent has a grounding norm that killing is wrong. In such an instance, she is disposed to feel guilty if she kills or anger if she hears of another killing. Now if this agent (upon reflection and drawing on

whatever non-emotional sources) determines that abortion falls under the category of killing, her reflective judgment that abortion is wrong is still emotion-based in the sense that (a) her anger at abortion or guilt at procuring one underlies this judgment or (b) she would be disposed to feel anger at others who procures an abortion or feel guilty if she procures one.

Now what about the possibility of the agent determining upon reflection that abortion is not wrong on the basis that abortion does not fall under the category of wrongful killing? Would such a reflective judgment be emotion-based (especially in light of the agent's lack of disposition to feel anger/guilt at abortion)? I think such a judgment can still be emotion-based, albeit in an indirect way. The agent's judgment that abortion is not wrong reveals her recognition (at least implicitly) that (a) abortion should not be prohibited as it is *not the sort of activity that would trigger her anger or guilt*, and (b) *she would have been disposed to feel anger or guilt* if abortion fell under the category of wrongful killing. It is in this sense that the agent's judgment that abortion is not wrong is emotion-based. Her reflective judgment (at least implicitly) refers to the disposition to feel anger/guilt under certain circumstances. Furthermore, the agent (in light of her judgment that abortion is not wrong) may be disposed to feel anger at those who attempt to restrict abortion or guilt if she does so. Her judgment that abortion is not wrong can remain emotion-based in the sense that it can be part of a network of other judgments like prohibiting abortion is wrong, restricting access to abortion is wrong, etc. And the latter set of judgments would still involve a disposition to feel anger/guilt.

A similar explanation can apply to an adjudicator's reflection on whether abortion is wrong. Suppose the adjudicator, disputants *A* and *B* shares the same grounding norm (that killing is wrong). Agent *A*'s position is that abortion is wrong. And as for agent *B*, abortion is not wrong. The adjudicator would have to resolve the dispute. As all three parties agree on the grounding norm, the key issue for the adjudicator to resolve can be whether abortion counts as killing. If yes, abortion is wrong. If not, abortion is not wrong.⁶⁶ Considering the preceding two paragraphs, the

⁶⁶ This example is admittedly a simplistic one in that it assumes that there is only one grounding norm involved and all are in agreement regarding this norm. It is with this assumption in mind that I claim that the key issue for the adjudicator to resolve is whether abortion counts as killing. If we admit other grounding norms, say being disrespectful is wrong, there may be other issues that the adjudicator would have to consider. The example's simplicity is mainly intended to accentuate how an adjudicator's reflective judgments can remain emotion-based.

adjudicator's reflective judgment would be emotion-based regardless of which side she takes. And if the adjudicator (for whatever reason) does not share the grounding norm with the agents, her reflective judgment can still be emotion-based. She may judge that abortion is wrong as the act falls within the category of another grounding norm. Or she may judge that abortion is not wrong in that it is not the sort of action that she would be disposed to get angry at/feel guilty about. And she may be disposed to feel angry at those who restrict abortion or feel guilty if she were to do so.

Emotion-based reflective process. With the above, we can see that an agent/adjudicator can draw on non-emotional sources within the reflective process. And judgment(s) that result from reflection can still remain in one way or another emotion-based. I turn the reader's attention now to the claim that the reflective process itself can also be emotion-based. This is the last and fourth step referred to at the start of this section. I start with the Mencian account. When King Xuan (in 1A7) asked can I care for my people, Mencius encouraged him to reflect on the compassion he felt that led him to spare the ox. Mencius was perhaps hoping that King Xuan could see the relevant similarities between the suffering of the ox and of his people. And if the king can be reminded of his compassion toward the ox, perhaps he can also see that he can care for his people. This reflection enabled the king to determine whether he can care for his people. What I would like to tease out from this narrative in 1A7 is a more general point that an agent when reflecting on moral questions can draw on emotional experiences. We can understand this general point from three vantage points. First, the moral questions involved can be broader than what is portrayed in 1A7. The agent can go beyond asking questions like can I care for x , to should I care for x , is caring for x wrong, etc. Second, the emotional experiences that the agent can draw on need not be limited to compassion. The agent can draw on other emotions within the reflective process. Third, the agent need not draw on only past emotional experiences. She can draw on the emotions felt when she focuses on different aspects of the situation. She can also draw on hypothetical ones, e.g. how would I feel about y if z . Now as I have explained in the preceding paragraphs how an agent can reflect on moral questions like should I see x as offensive, is abortion wrong, etc. (the first vantage point), I will not dwell on this further. Instead, allow me to elaborate on the second and third vantage points.

It is worth noting first that agents can not only draw on emotions within the reflective process, empirical studies suggest that emotions also facilitate this process. Allow me to begin with the claim that there appears to be a correlation between emotional incapacity and impaired personal/social reasoning. Damasio's observations of Elliot exhibits this point. Due to a brain tumour and its subsequent removal, damage was caused to Elliot's prefrontal cortices. A battery of tests revealed that no real damage was done to his intellect. The tests appear to reveal that he had "normal logical competence, normal attention, and normal working memory" (Damasio 1995, 43). However, subsequent observations revealed that Elliot could "recount the tragedy of his life with a detachment that was out of step with the magnitude of the events. He was always controlled, always describing scenes as a dispassionate, uninvolved spectator. Nowhere was there a sense of his own suffering, even though he was the protagonist" (Damasio 1995, 44). Elliot was subsequently asked to view emotionally charged images of death and destruction, but he admitted he no longer had any emotional response for such events. In other words and as it appears, the tumour damaged his emotional but not his cognitive abilities. This being said, Elliot's ability to make decisions in his personal and social life was severely impaired. His time-management skills suffered, he could not prioritize tasks in order of their importance/urgency and though he had business acumen, he had business ventures with disreputable characters which led to his own bankruptcy. It is this correlation that led Damasio to argue that emotions frame the reasoning process, i.e. the somatic marker hypothesis (Damasio 1995, 166-201).⁶⁷ Emotions mark information that the agent sees as important and such information would stand out in the reasoning process. Without such marks, the agent would appear to be flooded with information and would find it difficult to rank them in order and/or discard information that are unimportant. This can result in a poor personal/social judgement or an inability to come to such a judgment.

In sum, Damasio's findings suggest that when an agent engages in personal/social reasoning, her emotions can make salient and rank certain points of consideration. And if an agent suffers from an emotional deficit, she may not be able to properly engage in personal/social reasoning. While

⁶⁷ It is worth noting that Damasio's somatic marker hypothesis remains prominent in the literature of empirical psychology. I provide two examples. First, Martin Reimann and Antoine Bechara argue that "[d]espite some debate, the somatic marker framework is still providing a unique neuroanatomical and cognitive framework that helps explain the role of emotion in decision making" (Reimann and Bechara 2010, 767). Second, Vegard Olsen et al argue that "the somatic marker model of addiction contributes a plausible account of the underlying neurobiology of decision-making deficits in addictive disorders that is supported by the current neuroimaging and behavioral evidence" (Olsen, Lugo and Sütterlin 2015, 187).

Damasio's findings does not speak specifically to moral reflection per se, this need not impede us from using them to support the claim that emotions facilitate the process of moral reflection. And that without the emotions, an agent's ability to engage in moral reflection is likely to be impaired. As we can see, the gist of Damasio's finding is that emotions can mark certain information as important. The agent with an emotional deficit may not be able to identify what points of consideration is important (for example) when considering whether abortion is wrong. Or she may have too many points of consideration that would result in her inability to answer the moral question. Alternatively, she may not be able to identify which point(s) of consideration outrank(s) others. The take home message here is that Damasio's findings seem to suggest that the proper functioning of the moral reflective process would require the emotions.

The above hints toward emotions' indispensable role in the proper functioning of the reflective process, I now turn to how an agent can draw on a variety of emotional experiences within this process. I start with Goldie. He argues that "... we should respect our emotional responses, and listen to what they have to say to us and about us". On my reading of Goldie, our respect for emotional responses is on the basis that "our emotional responses can reveal to us what we value, and what we value might not be epistemically accessible to us if we did not have such responses..." (Goldie 2000, 48-49). In a similar vein, Michael Brady argues that "... emotion has epistemic value in virtue of directing our attention, quickly and efficiently, to objects and events that are of potential significance... emotion alerts us... to the importance of things that are suitably related to our cares and concerns" (Brady 2013, 25). This point that emotion reveals what we value is (at least on a high level) in line with Zagzebski's argument that "emotion reveals reality", that emotions detect "aspects of reality" and expressed in thick affective concepts (Zagzebski 2004, 77, 78). For Zagzebski, emotions reveal what we see as pitiable, offensive, disgusting, shameful, etc. And for Goldie, they can reveal what agents care about. When an agent is angry at someone who spews offensive remarks at her family members. This is likely a revelation that she cares for her family. We can now put the above within the context of a reflective process. Haidt argues that "[i]n the course of thinking about a situation a person may spontaneously activate a new intuition that contradicts the initial intuitive judgment... Simply by putting oneself in the shoes of another person, one may instantly feel pain, sympathy, or other vicarious emotional responses" (Haidt 2001, 819). The point to tease out here is reflection allows the agent to take on different perspectives. But while Haidt's point focuses on how reflection can take on the emotional perspective of *others*,

my focus would be on how reflection can allow the agent to focus on different aspects of a situation and on how by shifting her focus from one aspect to another can elicit differing emotional responses. The agent can draw on such emotional responses while considering moral questions. Adam Smith argues that a palace may be agreeable to the agent in its immediate effects (as we enter into the gaiety and comfort of the people who live there) but disagreeable in its remote effects (as it may promote distasteful greed and gluttony). On the flipside, a jail may be disagreeable in its immediate effects (as we enter into the suffering of those in it) but agreeable in its remote effects (as it provides a sense of security by keeping criminals out of the streets) (A. Smith 1790, I.II.24). While his focus is on how empathy can induce different emotions when the agent shifts her focus from the immediate to remote effects of the same situation, my focus is on a more general point, i.e. how differing emotions can be elicited when the agent's mental attention hones in on different aspects of the same situation. The agent can reassess her initial seeing of x as wrong/offensive/disgusting (for example) upon considering such emotional responses. As Bert Molewijk et al argue, emotions are relevant in ethics as "[e]motions may urge us to investigate whether the way we express the emotion is morally right, justified and balanced according to characteristics of the person and the facts of the situation" (Molewijk, Kleinlugtenbelt and Widdershoven 2011, 387). In addition to the point made here, a reflective inquiry can urge the agent not only to consider whether her emotional *expression* is appropriate/justified but also whether *being in an emotional state* is appropriate/justified. Allow me to consider an example.

When an agent sees late term abortion as wrong/offensive/disgusting, her mental attention may be focused on how such an abortion is procured. But the agent can on reflection shift her mental attention to various other aspects of the situation. Suppose the mother is only 12 years old. The agent may consider that the mother (being a child herself) is just too young to have a child. The agent may contemplate the difficulties that the mother would face in bringing up the child and feel compassion for her. Suppose as well that the mother's pregnancy was a result of rape. The agent may also in such a case feel compassion toward the victim, and an abortion may be thought of as a way of alleviating her suffering. The agent may also elicit anger, disgust, and/or contempt toward the rapist upon finding out that he is a family member of the victim. These emotions that reveal what the agent cares about may lean her toward seeing abortion as not wrong in this instance. This may change if the agent finds out that the mother wants to keep the child and it is the rapist that insists on the abortion. The agent may in such an instance see the rapist's insistence

as offensive. This may lean her toward seeing insisting abortion against the will of the mother as wrong. The agent can also approach the question of whether abortion is wrong from a hypothetical perspective. She can engage in a series of *what ifs* that can elicit different emotional responses and reveal what she values. This hypothetical reflective exercise can alter the agent's seeing of late term abortion as wrong/offensive/disgusting. With the above in mind, we can see that the agent may upon reflection have to decide among competing emotional responses (hypothetical or actual) in her attempt to answer the question whether abortion is wrong. The point of this example, of course, is not to resolve the question whether abortion is wrong but to illustrate that when an agent reflects on this question she can draw on a variety of emotional experiences. While the agent can draw on non-emotional sources within the reflective process, she is not precluded from drawing on emotional experiences. It is in this sense that the reflective process can remain emotion-based.

Summary

With the above, we see that at least some EAMs can account for agents' appeal to reflection and a *wide-range* of non-emotional sources to resolve moral qualms and/or disputes. More importantly, such EAMs can do so while retaining the position that reflective judgments can remain emotion-based. If reflective judgments can remain emotion-based, the fundamental issue that underlie the objection that emotion-based moral judgments are subject to something non-emotional appears to be resolved; the worry that agents' appeal to reflection and non-emotional sources as the ultimate arbiter undermines the essential involvement of emotions in moral judgments appears to be assuaged. This response to the adjudication problem can also (as a by-product) count as a reason in favour of EAMs as plausible accounts of moral judgment in the sense that it can account for the common phenomena of agents appealing to non-emotional sources to resolve moral questions. Furthermore, we also saw how the reflective process involved can also remain emotion-based. More specifically, the agent (while reflecting whether x has a moral attribute) can draw on her emotional experiences to resolve moral qualms and/or disputes. As such, moral judgment as *an activity*, as a process of reflecting on whether x has a moral attribute can in this sense be emotion-based. This too can count as a reason in favour of EAMs as plausible accounts of moral judgment as it accounts for a phenomenological experience common to many, i.e. when agents reflect on moral questions, it would not be uncommon for them to experience differing or even competing emotions.

Possible Answers Thesis

These reasons in favour of EAMs nicely segue us into a discussion on another reason in favour of EAMs as plausible accounts of moral judgment, i.e. for EAMs, there can be more than one possible answer to a moral question. As a shorthand, I will call this the possible answers thesis. We can consider this as a reason in favour in at least one way, i.e. we can understand the claim as a step away from parochialism and this step away can yield at least one benefit. Allow me to explain. Flanagan argues that “expanding inquiry beyond the resources of one’s own tradition and upbringing is confidence undermining in a worrisome way only if one has been encouraged to believe that ‘we’ have things more-or-less nailed down and other cultures are in various ways primitive, confused, immature, lost” (Flanagan 2016, 21). I see this as a criticism of the parochialism of academic moral philosophy as well as an invitation to broaden the epistemic base of moral inquiry. Flanagan’s argument continues from Alasdair MacIntyre’s: “the study of moral philosophy has become divorced from the study of morality or rather moralities and by so doing has distanced itself from practice” (MacIntyre 2013, 31) and Darwall et al’s: “[t]oo many moral philosophers... on moral philosophy... have been content to invent their psychology or anthropology from scratch and do their history on the strength of selective readings of texts rather than more comprehensive research into contexts” (Darwall, Gibbard and Railton 1992, 188-189). I do not intend to argue for this point as this goes beyond the scope of my thesis. Instead, I take it for granted. As such, the possible answers thesis would be a reason in favour of EAMs if the reader is inclined to accept the arguments of Flanagan, MacIntyre, and/or Darwall et al. In this light, the claim that there can be more than one possible answer to a moral question invites the reader to be open to the answers of those who are not ‘us’ as possibly right and possibly applicable not only to them but also to us. It invites the reader to be more attentive toward and less dismissive of differing moral judgments.

These invitations are wide open considering chapter four which outlined a variety of moral outlooks across and within communities and how such outlooks converge and diverge. And as Flanagan argues, “[t]hick description and charitable explanation of diverse moral worlds might make us more tolerant of the varieties of moral personality by helping us see why people are as they are” (Flanagan 2016, 20). In our context, we can be more tolerant of the moral judgments of

the 'other' by delving into how (a) agents across communities have shared emotions that (in one way or another) ground moral judgments and (b) such judgments can diverge in light of their local ecologies. More importantly, one possible benefit that can follow from this is that agents/adjudicators who eschew parochialism could likely engage in moral reflection in a more comprehensive and informed manner. In contemplating moral questions, the agent/adjudicator can not only draw on sources familiar to her, she can also draw on sources alien to her, sources that she normally would disagree with to determine an answer. While Flanagan, MacIntyre and Darwall et al focus mainly on how many academic moral philosophers are "imprisoned by [their] upbringing" (MacIntyre 2013, 31) and how they should break this mould, my focus is on how at least some EAMs can encourage agents/adjudicators to be less parochial while reflecting on moral questions that face them. In sum, if the reader sees that agents/adjudicators who are less parochial in moral reflection is a good thing, the reader can also see why the EAM claim that there can be more than one possible right answer can be a reason in favour of EAMs.

With the preceding paragraph in mind, the possible answers thesis is what I would focus on for the rest of this section. I start with Flanagan. He argues that "the right unit of attention for ethics is the whole person-in-communal relations, not person parts, say genes, or the emotional centers of the brain, or the rational parts of brains, not brains, period, but persons who seek to live well in relations with other persons in particular natural and social ecologies with histories" (Flanagan 2016, 21). We saw how this point can pan out in chapter four where Flanagan distinguished between first and second natures. Our first nature would be our initial emotional dispositions. And our second would be the result of our first nature's interaction with our ecologies. In my reading, this allows Flanagan to argue as well that "shared human nature is insufficient for flourishing and vastly underdetermines the possibility space for human lives" (Flanagan 2016, 21). While humans share some emotional dispositions (first nature), these dispositions by themselves rarely provide straight-up answers to our moral questions. Take shame for example. As argued in chapter four, an agent's shame directed at her personal inadequacy/weakness can be her first nature. She may also feel shame when she plagiarises. This can be her second nature (a disposition developed post-interaction with her local ecologies). But other agents (in different local ecologies) may not see plagiarism as shameful. This can give rise to the moral question whether plagiarism is wrong/shameful. We can see here that an agent's/adjudicator's inquiry into our shared first nature by itself would not likely put her in contact with the possible answers to this question. But

she could likely do so, when she examines the differing second natures. Possible answers to such a question are likely to emerge from our second natures. And the answers emerging from our second natures would likely vary as our local ecologies vary. While reflecting on whether plagiarism is shameful/wrong, the agent/adjudicator may come to see that plagiarism can be seen as (not) shameful for myriad reasons by agents in differing ecologies.

What I would like to tease out of the above is that while our shared emotional dispositions might constrain the possible answers to moral questions, there can still be a wide range of answers. Allow me to delve into this point by borrowing an argument from Nussbaum. She argues that humans have certain shared “grounding experiences” (e.g. facing the problem of resource scarcity). And the nominal definition of virtue is whatever the appropriate disposition to respond would be in that grounding experience (e.g. justice). The “job of ethical theory will be to search for the best further specification corresponding to this nominal definition, and to produce a full definition” (Nussbaum 1988, 36). In terms of how we can arrive at this ‘full definition’, Nussbaum argues in part that “the general answer to a ‘What is X?’ question in any sphere may well be susceptible of several or even many concrete specifications, in connection with other local practices and local conditions” (Nussbaum 1988, 44). What I take from her argument is that while humans share (a) a grounding experience (e.g. the question of how should we (re-)distribute our limited resources) and (b) ‘thin’ answer to how to deal with the experience (e.g. giving persons what they deserve), they may have different ‘thick’ answers to this question in light of their local ecologies. A Rawlsian’s thick answer may be to require a redistributive tax on the rich for the poor. The rich can keep their assets, but only if they contribute (by way of taxes) toward the plight of the poor. A Nozickian might, of course, reject this answer as such taxation is regarded as equivalent to theft. To the contrary, the Nozickian’s thick answer may be to require governmental institutions to respect the property rights of the rich and rely on philanthropy to alleviate the plight of the poor. At this stage, it would seem that Nussbaum’s agent would assess such and other competing thick answers to determine the ‘correct fuller specification’ of justice. Competing thick answers can be deemed right/wrong by assessing them against the standard of an “objective human morality based upon the idea of virtuous action – that is, of appropriate functioning in each human sphere” (Nussbaum 1988, 39). While the agent’s assessment may eliminate thick answers that are wrong, the agent may still conclude that (in light of differing ecologies) there may be more than one *right* thick answer to the moral question *how should we (re-)distribute our limited resources*. As Nussbaum

argues, “[w]hat is given in experience across groups is only the *ground* of virtuous action, the circumstances of life to which virtuous action is an appropriate response. Even if these grounding experiences are shared, that does not tell us that there will be a shared [thick] appropriate response” (Nussbaum 1988, 40). As a shorthand, I shall refer to Nussbaum’s argument here as the right answers thesis.

The EAM model can take on an approach similar to that of Nussbaum. While Nussbaum’s right answers thesis uses shared grounding experiences as a starting point, the possible answers thesis uses shared emotions. Here is an example. Nussbaum’s position is that humans across communities share the experience/problem of distributing limited resources. In a similar vein, the EAM model proposes, *inter alia*, that humans across communities share the emotion of anger toward the undeserved resource acquisitions by others.⁶⁸ Nussbaum’s position is also that humans across communities share a generic response to the experience/problem of limited resources, i.e. giving one what one deserves. Similarly, the EAM model proposes that humans across communities share a generic response to undeserved resource acquisitions by others (whatever form that such undeserved acquisitions might take). Via our shared anger, humans across communities share the tendency to respond with some form of punishment toward such undeserved acquisitions. It is at this stage that the right answers thesis diverges from Nussbaum’s thesis. Nussbaum’s generic response (i.e. giving one what one deserves) is a *thin* response to a *generic* problem of resource distribution. However, the EAM model sees its generic punitive response (motivated by anger) as an *extrapolation* of *thick* responses to *specific* problems of resource distribution.

Furthermore, while the thin response for Nussbaum can be *more fully specified* in differing ways in light of differing ecologies, the extrapolated response for the EAM model reveals how anger can be *tuned* in differing ways in light of differing ecologies. As argued in chapter four, conventions can tune an agent’s anger. More precisely, conventions can play a role in priming agents to construe specific events as *undeserved resource acquisition by another*. On one hand, Nozickian-like conventions may prime agents to see a capital gains tax on the rich as offensive: the government

⁶⁸ See ultimatum game experiment in chapter four.

has offended us by imposing a capital gains tax and such a tax (equivalent to forced labour) is bad; the government ought to be voted out/toppled. On the other, Rawlsian-like conventions may prime agents to see the refusal to implement a capital gains tax as offensive: the government has offended us by giving the rich excessive tax breaks and such tax breaks (being unfair) are bad; the government ought to be voted out/toppled. These, of course, are not the only ways in which anger can be tuned. Local ecologies (with differing conventions) would tune anger differently. What is likely to result from this is a range of possible answers to the problem of resource scarcity. The EAM model, in this sense, adopts an ecumenical approach. It aims at inviting one to be less parochial while reflecting on moral questions and (while it does not involve the claim that any of these possible answers are right) it does not preempt one from making such a claim. The agent/adjudicator aware of these answers may (upon reflection) determine that both the Nozickian and Rawlsian agents' anger are fitting/right as both agents are motivated by their concern for the miseries of the middle class. The agent's/adjudicator's determination may be based on her understanding that the difference between the Nozickian and Rawlsian lies mainly in their methods and that modelling predictions suggest that their methods may yield similar results in alleviating the miseries of the middle class.

With the above, we see that the EAM model serves at least two purposes. First, it can explain moral unanimity and diversity. Second, the model draws attention to how emotion-based judgments can be tuned in different ways. These two points accentuate how there can be more than one possible answer to a moral question. Emotion-based judgments (whether or not reflected upon by agents) can vary across and within communities. And what can follow from this is that answers to moral questions can vary across and within communities. The EAM model draws our attention to this phenomenon. And it is in this sense that the EAM model can enhance an agent's/adjudicator's awareness of what Taylor argues to be an "ethnocentric temptation... to make too quick sense of the stranger, i.e., sense in one's own terms" (Taylor 2003, 138).

Concluding Remarks

With the above, we also see another reason in favour of EAMs by way of the possible answers thesis. This thesis encourages agents to be anti-parochial while reflecting on moral questions. The

novelty of the possible answers thesis lies in how it was developed, i.e. how it use Nussbaum's insightful right answers thesis as a starting point and how the former distinguishes itself from the latter. With this in mind, we can now compare the EAM model's explanation of moral unanimity and diversity with a rival account.

Chapter Six

"I once heard that Zixia, Ziyou, and Zizhang each had one of the qualities of the sage, while Ran Niu, Min Zi, and Yan Yuan had all of the qualities, but in slighter degree. I venture to ask with which of these you are willing to be compared?" (Mencius c 300 BC: 2009, 2A2)

A Comparative Exercise

Allow me to recount the journey we have undertaken so far. In chapter one, we considered a variety of emotion theories and the assumptions I would carry into the subsequent chapters of my thesis. In chapter two, we sampled a variety of western emotion-based accounts of moral judgments. And in chapter three, we considered a Mencian account of moral judgment and how this account (especially with its use of metaphors) sets the scene to reveal how such emotion-based accounts can explain moral unanimity and diversity. In chapter four, we saw how an emotion-based approach to moral judgments can be apt in explaining moral unanimity and diversity. And in chapter five we saw how emotion-based accounts can incorporate reflection and appeals to non-emotional sources while retaining the essential relationship between emotions and moral judgments. We also saw how (a) the possible answers thesis can flow from the EAM explanatory model, (b) the thesis can enhance an agent's/adjudicator's awareness (while reflecting on moral questions) of the ethnocentric temptation, and (c) the thesis can serve as an additional reason for emotion-based accounts if the reader sees anti-parochialism as a good thing. The reader's attention has, as such, been resting largely on how human moralities could have begun with a common emotion-based starting point and how they could have subsequently diverged. She is also made aware in this context of the varying but possible answers to moral questions.

In this final chapter, I intend to compare the EAM model with Gilbert Harman's conventionalism, illustrate how Harman's conventionalism (as a rival to the EAM model) can explain moral unanimity and diversity, and argue that Harman's conventionalism loses out in at least one respect when compared to the EAM model (i.e. while conventionalism would have to refer to emotion to

explain instances of moral unanimity, EAMs would not have to refer to agreements to explain such instances). The above is done with the aim of ending my thesis on the note that not only are there reasons in favour of an EAM model in explaining moral unanimity and diversity, the model can also best a rival explanatory account in at least one respect. Here also lies the novel contribution of this chapter. No detailed comparison (at least in the manner set out) has been done between EAMs and Harman's conventionalism (a prominent account of moral judgment) in their explanations of moral unanimity and diversity. I will return to this later. For now, allow me to begin with an outline of this rival account (and why I compare the EAM model with Harman's conventionalism).

(Why) Conventionalism

Inner judgments and agreement. Harman argues that "inner moral judgements are made in relation to an implicit agreement" (Harman 2000, 13). This, in a nutshell, is conventionalism. Before we consider in detail how inner judgments *relate to* an agreement, let us first consider what Harman takes to be inner moral judgments. We can divide our consideration into two questions. First, what Harman takes to be a *moral* judgment. Second, what he takes to be an *inner* judgment. Allow me to start with the first question. Harman argues that *moral* judgments include judgments "in which we say that someone morally should or ought to have done something or that someone was morally right or wrong to have done something" (Harman 2000, 4). In other words, a moral judgment includes judgments like *agent A should not have procured the abortion* and *agent A's procurement of an abortion is wrong*. With this in mind, let us consider what Harman takes to be an *inner* judgment. For him, *inner* judgments are a subset of *moral* judgments. He argues that inner judgments "imply that the agent has reasons to do something... [and] the speaker [of such judgments] in some sense endorses these reasons and supposes that the audience also endorses them" (Harman 2000, 8). Harman also argues that "... if S says that (morally) A ought to do D, there are certain motivational attitudes M which S assumes are shared by S, A, and S's audience" (Harman 2000, 10). Allow me to explain these points with the abortion example. The *moral* judgment *agent A should not have procured the abortion* or *agent A's procurement of an abortion is wrong* would be an *inner* judgment only when the speaker of this judgment (a) has the motivational attitude or endorses the reasons to avoid procuring an abortion, (b) assumes that agent A (is likely to have or) has the motivational attitude or (is likely to endorse or) endorses the reasons to avoid

procuring an abortion, and (c) assumes that *her audience* has the motivational attitude or endorses the reasons to avoid procuring an abortion. Conditions (a) to (c) can be satisfied when (for example) the speaker, agent and audience are staunch Catholics. These conditions can also be satisfied when the speaker is a staunch Catholic and the agents and audience, while not being Catholics are within “the motivational reach of the relevant moral considerations” (Harman 2000, 7). The agent and audience may be within this ‘motivational reach’ in the sense that they (in one way or another) respect the sanctity of life. All of them would (presumably) be capable of being motivated to avoid procuring an abortion.

At this point, the reader can see that an inner moral judgment can take the form of *A should not commit B* or *A’s commission of B is wrong*. With this in mind, let us probe deeper into the nature of such inner judgments. Inner judgments can be understood as a psychological state whose content can be represented by a proposition. Allow me to explain with a conundrum arising out of David Copp’s argument that “[i]nner judgments are a kind of moral judgment, and moral judgments are a kind of *statement* expressed by certain sentences in certain circumstances [emphasis added]” (Copp 1982, 229). The question now is how do we best make sense of Copp’s stance. On one hand, a statement or proposition by itself *does not* have a location. On the other, the term ‘inner’ (in inner judgments) suggests that they *do* have a location of sorts (psychological states are commonly regarded as *inner* states). Copp does not clarify this point. But I would attempt to do so. To reconcile these two positions, we can say that inner judgments are psychological states whose content can be represented by a statement/proposition. The part on inner judgments being psychological states is consistent with Copp’s view that we cannot (on the face of the statement itself) determine whether a judgment is an inner judgment: whether judgments are inner or noninner “has to do with what the *speakers* have implied, not with the content of the statements or judgments they have expressed” (Copp 1982, 232). We can see this as Copp’s attempt to ‘situate’ an inner judgment in the psychological state of the agent. And while we do so, we need not abandon the claim that there is a nexus between inner judgments and statements. The psychological state of the agent can, for example, be represented in the statement/proposition that *A’s commission of B is wrong*. Once represented as such, the statement/proposition can be true or false. This ‘squares’ also with Harman’s argument that “moral judgements *are true or false* only in

relation to... an agreement [emphasis added]" (Harman 2000, 19).⁶⁹ While psychological states cannot be true or false, propositions that are representations of psychological states can be.

The above take on Copp's position would also appear to be consistent with Prinz's view on inner judgments. The latter argues that the inner judgment *A ought not do x* can be understood as a "propositional oughtitude" (Prinz 2007, 175-176). For Prinz, an agent's psychological act of judging that *A ought not do x* is an oughtitude only when the agent herself is disposed to feel 'other-blame' emotions like anger when *A* commits *x* or 'self-blame' like guilt when the agent herself commits or contemplates the commission of *x* (Prinz 2007, 96, 175). As we can see, Prinz's understanding of inner judgments establishes a conceptual link between inner judgments and psychological states, and specifies what underlies the agent's motivational attitude in avoiding *x* or what it means for an agent to endorse reasons to avoid *x*. At this point, one may ask *how then can inner judgments (understood as propositional oughtitudes) be the kind of thing that can be true or false*. The answer to this again appears to be that such judgments (as psychological states) are *not* the kind of things that can be true or false. More precisely, the agent's state of judging that *A ought not do x* cannot be true or false. Rather it is the proposition that represents such an inner judgment that can be true or false relative to agreements.

Inner judgments, as such, would be in line with how some EAMs understand moral judgments. And if both conventionalism and (some) EAMs see moral judgments as psychological states of sorts, the reader need not be concerned that (at least in this respect) we would be comparing the 'oranges' of conventionalism with 'apples' of EAMs. While an in-depth discussion has been conducted in chapter two on how EAMs see moral judgments, here are two examples as a refresher. For Zagzebski, a moral judgment can be the *psychological state* of seeing *x* as having a thick affective property (plus assent) or seeing that *x* has a thick affective property (plus assent). These same states can be put in a proposition, i.e. *x is pitiable*. For Prinz, an agent's moral judgment can also be understood as a psychological state. More precisely, an agent's recognition that she has a disposition to feel 'other-blame' and 'self-blame' emotions when contemplating *x* or when *x* occurs (for example) can be understood as a moral judgment. This judgment can also be represented in

⁶⁹ As inner moral judgments are a subset of moral judgments, this point applies to inner moral judgments as well. In other words, inner moral judgments for Harman are true or false only in relation to an agreement.

the proposition *x is wrong*. The point here is that we can engage in a comparative exercise between Harman's conventionalist account of moral judgments with *at least* some emotion-based accounts of moral judgments.

With the above in mind, we can now examine how Harman's inner moral judgments relate to agreements. He argues that (a) "when a relevant speaker *S* makes the inner judgement that *A* ought to do *D*, *S* assumes that *A* intends to act in accordance with an agreement that *S* and *S*'s audience also intend to observe" and (b) such an agreement is the "result of some sort of bargaining, either explicit or implicit" (Harman 2000, 9, 12). Allow me to explain in more detail. Regarding (a) and in the context of our abortion example, the speaker expressing the judgment *A's procurement of abortion is wrong* assumes that herself and her audience intend to observe an agreement that the judgment/proposition relates to. The agreement that the judgment relates to can be one that was explicitly or implicitly entered into by members of the Catholic community. This can help explain the speaker's assumption (while expressing the judgment/proposition) that *A* would be motivated to comply with the agreement. The speaker would expect her fellow Catholic *A* to adhere to the teachings of the magisterium. As for (b), it allows us to see how such agreements could have been formed. We have interests and want to further them. We form "conditional intentions" to act in a particular manner as a result. We then hope that other agents would share such intentions. However, other agents may have interests different from ours and hence differing intentions. It is at this stage that bargaining (explicit or implicit) occurs and a compromise is reached between us and other agents. The result of this compromise is an agreement (Harman 2000, 12). In our context, the agreement that the judgment *A's procurement of abortion is wrong* relates to could have resulted from a bargaining process involving (perhaps) different Catholic factions (for example, Catholic conservatives and liberals). The bargaining process need not involve all parties that would be affected by the agreement. Involvement in the bargaining process may be limited to the Catholic elite (the pope, cardinals, bishops, and clergy), with the Catholic laity and religious sisters having little or no say. This is in line with Harman's position that "morality derives from an agreement among people of varying powers and resources" (Harman 2000, 11). Some may have more say in the bargaining process over others in coming to an agreement.

The inner judgment *A's procurement of abortion is wrong* could also relate to an *implicit* agreement. For Harman, an agreement need not be a ritual (where parties meet, exchange terms, negotiate, bargain, etc. prior to entering into a formal agreement). In fact, many (or even most) agreements that inner judgments relate to are implicit ones. There is no need (for example) for bishops to convene the Second Vatican Council, to debate and vote on the issue of abortion for there to be an agreement that abortion is wrong. As Harman argues on implicit agreements, "there is no single moment which one agrees, since one continues to agree in this sense as long as one continues to have the relevant intentions. Someone refuses to agree to the extent that he or she does not share these intentions" (Harman 2000, 16). Applying this point to our context, a Jew, Muslim, Hindu, Buddhist, and Catholic living in the same community may have an *implicit* agreement in the sense that each of them intends to refrain from procuring abortions (at least in certain circumstances) on the assumption that every other person intends to do the same. They have an agreement in the sense that all of them have the relevant intentions. And there is no need for any actual bargaining to have preceded such an agreement. Consider another example. The inner judgment *A should not steal* can relate to an implicit agreement that was not preceded by any actual bargaining. Instead, each party to the agreement could have a shared anger toward those who steal from them and, thereby, intend not to steal from others on the assumption that other parties to the agreement intend the same. Augustine of Hippo argues that "[w]ithout doubt, [God's] Law punishes theft... and so does the law which is written in the hearts of men, which even iniquity itself does not erase. For, what thief will suffer from another thief without protest? Not even he who has plenty when the other has stolen under the impulse of want" (Saint Augustine c 400AD: 1953, 39-40). Though Augustine probably intended this as an argument for a natural law against theft, his argument can be co-opted to suggest that humans across communities share an anger toward those who take what is theirs without permission and with no intention to return it; even thieves would be angry at those who steal from them. And if there is such a shared anger among a set of human individuals, this set of human individuals could have arrived at an agreement not to steal without the need for any actual bargaining.

With the above in mind, we can now start to see a key distinction between EAMs and conventionalism. As argued in chapter two, EAMs are united by the thesis that emotions are in one way or another *essentially involved* in moral judgments. As such, one way of distinguishing EAMs from conventionalism is to pose the following question to the conventionalist: *what are moral*

judgments essentially involved with. The conventionalist answer would likely be moral judgments are essentially involved with agreements. For the conventionalist, the term ‘wrong’ in the proposition *A’s procurement of abortion is wrong* relates to agreements. The judgment *A’s procurement of abortion is wrong* is a shorthand for *A’s procurement of abortion is wrong according to an agreement*. An EAM theorist’s answer, by contrast, can be that moral judgments are essentially involved with emotions. Here we see conventionalism as a rival to EAMs in that the former offers an alternative story to the question of *what are moral judgments essentially involved with.*

The above nicely segues us into the question *why compare EAMs specifically with Harman’s conventionalism.* Allow me to briefly address this (as there are after all many other conventionalist or contractarian accounts of morality). My choice of Harman’s conventionalism is motivated mainly by three reasons. I have alluded to these reasons earlier in this section, but I now make them explicit. The first reason is that Harman’s account provides us with a convenient starting point to engage in a comparative exercise: just as EAMs outline accounts of moral judgment, Harman’s conventionalism explicitly and clearly outlines an account of moral judgment. His account is clear as Harman lays bare what he thinks to be the logical form of inner judgments. An inner judgment is understood as “involving a four-place relation. ‘Ought (*A*, *D*, *C*, *M*)’ relates an agent *A*, a type of act *D*, considerations *C*, and motivating attitudes *M*” (Harman 2000, 9). The proposition that represents the inner judgment *A’s procurement of abortion is wrong* can thus be more fully adumbrated in the follow manner: ‘Considering that *A* is a Catholic (*A* would presumably have the motivational attitude to avoid procuring an abortion), *A* ought not to have procured the abortion’. On a related point, Harman’s focus on *inner* judgments uniquely positions his account as a point of comparison for EAMs: just as many EAM theorists see moral judgments as psychological states (of one form or another), Harman also sees inner judgments as a kind of psychological state. Apart from this, the content of such inner judgments can also be represented by a proposition. This position is also close to the EAM position. For EAMs, moral judgments can be understood as (a) psychological states that can be verbalised and given a propositional form, (b) a kind of proposition, and/or (c) a kind of utterance. The second reason is that Harman’s conventionalism (like EAMs) is in the explanatory game. In chapter four, I argued that the EAM model can plausibly explain two salient features of morality, i.e. moral unanimity and diversity. I also argued that the EAM model’s explanatory power is a reason in favour of EAMs as plausible accounts of moral judgment. Similarly, a ‘selling point’ of Harman’s conventionalism (according

to him) is its ability to “explain otherwise puzzling aspects of some of our own moral views” (Harman 2000, vi). He argues, for example, that “the hypothesis that morality derives from an understanding among people of different powers and resources can explain why in our morality avoiding harm to others is taken to be more important than helping those who need help” (Harman 2000, 11). The rich and poor alike would benefit from the ‘harm no one’ rule. But the poor stands to gain more and the rich to lose more with the ‘help others always’ rule. Since everyone can agree to the former rule and not everyone can agree to the latter rule, this appears to be able to explain why the former rule is more important than the latter. Harman also argues that his account can easily explain “the difference in moral status of people and animals” (Harman 2000, 69). Agreements are arrived at by humans and for humans. Animals do not participate in the bargaining process. If a section of the human population disagrees with a rule, they can lobby for its abolishment. A compromise could then arise out of this lobbying. Animals obviously do not have this option. In addition to the above, conventionalism appears also to be promoting its ability to explain instances of moral unanimity and diversity. I will address this point with more detail in the following two sections. It would suffice to note for now that Harman (for example) argues that there may be a completely conventional explanation for our shared concern and respect for others.

The third reason is that conventionalism is a clear rival to EAMs. This ‘flows’ from the first and second reasons. With the above two reasons, we see that Harman’s conventionalism operates in the same ‘space’ with EAMs. Both offer accounts of moral judgment. Both speak of moral judgments in terms of psychological states. Both play the explanatory game. However, both are at odds with each other on what moral judgments are essentially involved with and their explanation of moral unanimity and diversity. This sets conventionalism up as a clear rival to EAMs. And as a clear rival, it would be apt to compare conventionalism with EAMs to see how they fare against each other. Without further ado, let us now proceed to examining how conventionalism would serve as a rival in the ‘space’ of explaining moral unanimity and diversity.

Conventionalism's Explanation of Moral Unanimity

As argued in chapter four, the EAM model can explain instances of moral unanimity by clustering them around specific emotions. Here are three instances. The first instance of moral unanimity would be our principles/conventions on alleviating suffering. Across communities, we share the principle that 'those in need should be cared for'. And communities have conventions that (in one way or another) care for those who are (in one way or another) in need. EAMs can explain these pancultural principles and conventions by having them clustered around compassion. The second instance of moral unanimity would be the principles and/or conventions on respecting our parents. I argued that across communities, we share the principle that 'parents should be respected'. And communities have conventions where children (in one way or another) respect their parents. EAMs can explain such principles and practices by clustering them around deferential feelings. The third instance would be the principles and/or conventions on justice. Across communities, we share the principle that 'one should be given what one deserves'. And communities have conventions that (in one way or another) mete out justice. EAMs can explain these principles and practices by clustering them around anger.

On conventionalism's part, Harman is willing to accept that there may be some level of moral unanimity. He argues that "[i]t may be true that rules against killing, harm, lying, and cheating occur in all societies" (Harman 2000, 218). Specific to principles/conventions on the alleviation of suffering and respect for parents, Harman argues that "[t]here is clearly some connection between morality [in general as opposed to a morality specific to a community] and concern and respect for others" (Harman 2000, 70). In entertaining the possibility that we are "genetically constructed so as to feel such concern and respect for others" (Harman 2000, 70), he appears to imply that there is some level of unanimity in our conventions involving concern and respect for others. As for principles/conventions on giving one what one deserves, he argues "there may be *universal* formal principles of justice [emphasis added]" (Harman 2000, 58). Considering this, conventionalism offers a rival (and non-emotional) explanation to the above instances of moral unanimity. Harman argues that there may be "a *completely conventional explanation* of our concern for others... [where] one develops concern and respect for others *as part of accepting* [an agreement], on the supposition that others are developing or have developed similar concern and respect [emphasis added]" (Harman 1996, 26). Allow me to consider this in detail.

On the conventions on alleviating suffering, the conventionalist can say that these can be explained with reference to agreements. Due to resource scarcity (food, water, land, etc.), members within a group may easily resort to violence to take what is needed. While the strong may not be so motivated to bargain with the weak, even the strongest must sleep: “[f]or as to the strength of the body, the weakest has strength enough to kill the strongest, either by secret machination, or by confederacy with others...” (Hobbes 1651: 1965, 94). This may motivate members within a group (despite their differing bargaining positions) to enter into negotiation. Members within this group can reach agreement in the sense that each member in this group intends to alleviate suffering on the understanding that other members of the group have such intentions as well. As Harman argues, “[t]here is an agreement... if each of a number of people intends to adhere to some schedule, plan, or set of principles, intending to do this on the understanding that the others similarly intend” (Harman 2000, 4). This agreement to alleviate suffering could hence have arisen to manage the problem of resource scarcity and avoid unnecessary violence. If members can agree to share their resources (to an extent) with those in need, there is less need for some members within a group to forcibly (and violently) take resources from other members. Such an agreement could have come about via an *explicit* bargaining process, where members of a group convene a roundtable discussion (for example). Alternatively, such an agreement could also have come about with no such explicit bargaining. The following is a hypothetical example on how such an implicit agreement could have arisen. Agents *A* and *B* could have been stranded on an island. Both may share the insight that if *A* helps *B* when *B* is in distress (and vice versa), *A* and *B* would have better odds at surviving on this island when compared to a situation where *A* and *B* do not help each other out when in distress. For example, *A* has a better chance of fending off dangerous wild animals with *B*’s help when *A* calls for help (and vice versa). This shared insight on *A*’s and *B*’s part, could have resulted in an agreement in the sense that it led to *A* developing an intention to help *B* when *B* is in distress with the understanding that *B* intends to do the same when *A* is in distress. The shared insight could also have led to *B* developing an intention to help *A* with the understanding that *A* intends to help *B*. As we can see, no explicit bargaining is needed for *A* and *B* to come to such an agreement. And if *A* and *B* find *C*, *D* and *E* (and so on) on this island, all of them could have the shared insight and subsequently come to an agreement. Once such kinds of agreements to alleviate suffering is established (whether explicitly or implicitly), members of the

group could have subsequently developed such a concern for others via practice and habituation (Harman 1996, 26).

As for conventions on respecting our parents, the conventionalist can similarly say that these can be explained with reference to agreements.⁷⁰ It is uncontroversial to say that minor children are highly dependent on their parents to take care of them and parenting is a difficult task. On top of this, it is also a very costly business (in several ways) for parents to take care of their children from infancy until adulthood. Members of a group (especially parents) could have interests in ensuring that parents are recognised for what they are doing or have done. Members of this group, however, may disagree on how children should respect their parents and what are the exceptions to it (if any). This could have led them into a bargaining of sorts. An agreement in the sense that some members of this community intend to ensure that children respect their parents with the understanding that some other members intend to do the same could have resulted from this bargaining process. Children may or may not have taken part in this process that led to the agreement. While minor children who can express their interests might have taken part in the bargaining process and they may or (at certain ages) would likely have interests that differ from their parents, their interests may not have been given much weight due to their weaker bargaining position. The phrase ‘if you live under my roof, you live by my rules’ succinctly addresses why children might be in a weaker bargaining position. Again, the agreement to respect parents need not result from any explicit bargaining process. Here is an example of how an implicit agreement to respect parents could have arisen. The devotion of parents in caring for their children could have motivated the latter’s intention to act in respectful ways toward the former. Stern looks by parents after their children ‘misbehave’ could also have had the same effect. The children’s intentions could also be coupled with an understanding on their part that their parents have cared for them and intend to continue to care for them. Similarly, parents could also have the intention to care for their children on the understanding that their children intend to act in respectful ways toward them. At this point, we see that an agreement to respect parents could have been ‘made’ without the need of any spoken word between parents and children. But regardless of whether

⁷⁰ The reader may have noticed that I shifted from a discussion on shared respect in general to shared respect of children toward parents. I note that while Harman’s discussion on shared respect appears to be a general one (as opposed to a specific shared respect of children toward their parents), I see no reason why Harman’s wider notion of shared respect cannot apply a more specific instance of shared respect. The former appears to encompass the latter.

such kinds of agreement to respect parents came about explicitly or implicitly, once the agreement is in place, the children's show of respect (in one way or another) toward their parents can be developed via practice and habituation.

For the conventionalist, conventions on giving one what one deserves can also be explained with reference to agreements. The rich and poor alike may have an interest in escaping any liability and living their lives with impunity. This calls to mind Glaucon's question in Plato's *Republic* on whether one would ever act justly if given the ring of Gyges (making one invisible and able to avoid any detection). While the rich and poor may share such an interest, they may also have other interests. The rich may want to keep their prized possessions safe, while the poor may want to keep themselves safe from the excesses of the rich. They may as a result enter into a bargaining process and come to an agreement to punish wrongdoers for their wrongdoing (retributive justice). As Harman argues, "... the basic principles of justice accepted by people of different powers and resources are the result of a continually changing compromise..." (Harman 2000, 61). Specifically and in our context, each rich or poor person may intend to punish wrongdoers for their wrongdoing with the understanding that the other rich and poor intend to do the same. Such an agreement to punish wrongdoers could have also arisen with no explicit bargaining process. Harman argues that "[i]t may be that everyone has some innate inclination towards 'fairness', but it seems likely that, in the absence of actual bargaining, agreement is possible only on a formal conception of justice, not a substantive conception" (Harman 2000, 61). What I want to tease out of this is that Harman seems to suggest the possibility of arriving at an agreement on formal principles of justice without the need for any *actual* bargaining. In our context, the formal principle would be 'wrongdoers should be punished'. It is a formal (as opposed to substantive) principle in the sense that it leaves out details on what counts as wrongdoers/wrongdoing and what the appropriate measure of punishment would be. An agreement to punish wrongdoers in this formal sense could have arisen out of an unspoken interest and shared inclination of members of a community to punish wrongdoers: 'don't mess with me/us, or else, you'd be very sorry'.

With the above, we see how (some instances of) moral unanimity can be explained with reference to agreements. At this juncture, it would be apt for me to consider a plausible conventionalist answer to the question *how innumerable (explicit or implicit) bargaining processes across communities*

could have led to a convergence of such agreements. Harman's conventionalism can be aided with an argument from Nussbaum: there are "sphere[s] of human experience that [figure] in more or less any human life, and in which more or less any human being will have to make some choices rather than others, and act in some way rather than some other" (Nussbaum 1988, 35). These spheres of experience include the following:

- (1) We fear "important damages", especially death.
- (2) We face problems on the "[d]istribution of limited resources".
- (3) We face "slights and damages" caused by other individuals (Nussbaum 1988, 35).

Nussbaum argues that each of these spheres of experience has a corresponding virtue (for example, courage is the corresponding virtue when we fear damages to self). This may be the case. But my use of her argument serves instead as a precursor to showing how there can be a convergence of certain kinds of agreements across communities. Nussbaum's argument does just that by revealing how certain questions (and not others) could have arisen due to our shared environmental conditions. Our fear of important damages to self or property raises questions like *how should we deal with threats to ourselves and property*. Questions on how should we deal with threats to our physical self (for example) would not likely have arisen if we were physically invulnerable. Resource scarcity raises questions like *how should we distribute such limited resources*. Such a question would likely not have arisen if we had infinite resources. And slights and damages caused by others raise questions like *how should we deal with those who slight us or cause us damage*. Questions of this sort may be reduced if we are not the sort of creatures susceptible to slights.

With Nussbaum's spheres of experience, we see how certain common features of human life can raise certain common (kinds of) questions. This nicely segues us into a discussion on how "basic evaluative tendencies" (possessed by and endowed unto us humans by selective pressures) (Street 2006) can narrow down the ways we respond to such questions. Sharon Street argues that the list of tendencies include the following:

- (1) "The fact that something would promote one's survival is a reason in favor of it".
- (2) "The fact that someone has treated one well is a reason to treat that person well in return".
- (3) "The fact that someone has done one deliberate harm is a reason to shun that person or seek his or her punishment" (Street 2006, 115).

As we can see, Street frames these tendencies in terms of the things that humans have reasons to do. However, it would suffice for my purposes to recognize that (if Street is correct), when faced with certain questions in certain shared spheres of experience, we are disposed to promote our own survival, promote reciprocal relations, and punish wrongdoers.

At this stage, we are able to apply what is extracted from Nussbaum's and Street's arguments to the question *how innumerable (explicit or implicit) bargaining processes across communities could have led to a convergence of such agreements*. I will do so with the kinds of agreements on alleviating suffering and punishing wrongdoers. Allow me to start with the former. When faced with the question *how should we deal with threats to physical integrity*, we may be disposed to promote our own survival and promote reciprocal relations. We may (as a result of or alongside such dispositions) come to recognize that our odds at surviving threats to our physical integrity increases when we help each other out. While some members who (for whatever reasons) are less susceptible to such threats may be less inclined to answer *help them out*, these members may also recognise the 'insurance policy' of others having their backs. If I help them out now, they would be more inclined to help me out in the future. As we can see, our basic dispositions to promote our own survival and reciprocal relations coupled with the recognition that "*mutual cooperation* [consisting of individuals benefiting others in return for some future benefit] is better for everyone involved than *mutual defection* [consisting of individuals refusing to benefit others in return for some further benefit]" (S. M. James 2011, 42) could have narrowed down the types of answers we would give to the question *how should we deal with threats to physical integrity* and left us with answers along the lines of *we would help them out on the understanding that they would also intend to do the same if we were in their situation*. On agreements on punishing wrongdoers, when faced with questions of the sort of *how should we deal with those who take without permission crops grown in our farms*, we may be disposed to shun or punish such people. The ultimatum game (discussed in chapter four) also highlights this strength of our desire to punish a perceived unfairness. It is clear from such experiments that "this sense of fairness can drive people to punish others – even if it costs them personally" (S. M. James 2011, 68). With this in mind, it appears that we would even prefer to punish the perceived wrongdoer and lose out on what we could have gained (if we chose not to punish). Such a disposition to punish coupled with our recognition that resources are limited and the hard work that we had likely put into growing the crops would exclude answers along the lines of *let them take whatever they want*. Instead, our shared recognition on resource

scarcity, our shared resentment at those who take without permission what we have worked hard for and our shared retributive urge could have left us with answers along the lines of *we would punish those who take without permission on the understanding that others would intend to do the same.*

Conventionalism's Explanation of Moral Diversity

With the conventionalist explanation of moral unanimity discussed, I can turn to the task of showing how conventionalism can explain instances of moral diversity. As a refresher, I argued in chapter four that while there may be some level of moral unanimity in that (for example) all communities in one way or another restrict the killing of human beings and sexual activities, such prohibitions pan out in very different ways across and within communities. In terms of the 'no killing' rule, some communities (for example) allow honour killings as an exception to this rule, but others do not. As for the restriction of sexual activities, some communities (for example) see homosexual and/or pre-marital sexual intercourse as permissible, but others do not. These examples suggest that what appears to be the smooth surface of moral unanimity has pockets of moral diversity.

Now let us consider the conventionalist position on moral diversity. Harman argues that "[i]t may be true that rules against killing, harm, lying, and cheating occur in all societies. But there is a sense in which these are not exactly the same rules, since the protected group changes from one society to another" (Harman 2000, 218). Allow me to narrow my focus down to the 'no killing' rule mentioned. While Harman is ready to accept that there is some level of unanimity across communities regarding this rule, he also argues that the 'no killing' rule of one community would differ from another if we probe deeper. The difference can lie in the designated group(s) of each community that are exempt from the rule. Harman argues that "[i]n many societies there are no limitations on what a husband can do to his wife or on what a father can do to his young children" (Harman 2000, 218). Here the honour killing exception comes to mind. The group exempt from the 'no killing' rule in this instance would be that of husbands/fathers/brothers. Harman's point can also be extended to sexual restrictions across communities. While all communities may have the 'no rape' rule in one form or another, the exceptions to this rule vary from community to community. In some instances, the exception to the rule can be understood within the context of

how rape is defined. Here the exception of marital (nonviolent) rape comes to mind. While some communities see marital rape as wrong, others do not even recognise that rape can occur within a marriage. In communities of the latter sort (e.g. Saudi Arabia), non-consensual intercourse within a marriage can be understood as an exception to the 'no rape' rule. Applying Harman's point to this context, husbands may constitute the group that is exempt from the 'no rape' rule in such communities.

In addition, just as I have argued in chapter four that moral diversity exists across and within communities, Harman argues that "[t]here are two kinds of apparent moral diversity, *societal and individual* [emphasis added]" (Harman 2000, 217). The former refers to *inter-community* diversity. Such diversity ranges from differences in etiquette (e.g. whether burping after meals is polite) to differences in bigger issues like whether cannibalism or slavery is permissible. And the latter refers to *intra-community* diversity. As Harman argues diversity "occurs not just between societies but also within societies, and in a way that often leads to seemingly intractable moral disagreements that rest on irreconcilable differences in basic values and not just on differences in opinion about the nonmoral facts" (Harman 2000, 218). One example that comes to mind here is the debate whether there should be stricter gun control laws in the USA. If we consider the USA as one community, we can understand the tension between those who favour stricter gun control laws and those who do not as an instance of intra-community diversity. Both movements are at odds with each other and their disagreement can rest on differences on the prioritization of values (e.g. the liberty of a person to carry arms) as opposed to any differences on nonmoral facts (e.g. whether there has been an increase in gun violence).

As we can see, the conventionalist position on moral diversity is similar to that of EAMs. With this in mind, we can now compare the EAM explanation of (instances of) moral diversity with that of the conventionalist explanation. We can briefly summarise the EAM model (as argued in chapter four) as follows: different *emotion-based* conventions can tune emotions/emotion-based moral judgments. This tuning process can explain why agents of Community *A* have/make certain moral judgments while agents of Community *B* do not have/make these judgments. The presence of such judgments within communities can sustain and/or lapse existing conventions within those communities and can also result in the generation of new conventions within those communities.

The gradual absence of certain moral judgments within communities can result in the lapsing of existing conventions. In short, moral diversity (the presence/absence of certain conventions among communities) can be explained with reference to *emotion*.

To allow us to compare the EAM model with the conventionalist model, let us consider one instance of moral diversity. Community *A* may have the convention of refraining from eating factory farmed animals while Community *B* may not. The EAM model can explain this instance of diversity in the following way: While both communities have agents who feel compassion (toward one thing or another), there are insufficient or no agents in Community *B* that are disposed to feel compassion toward factory-farmed animals. The absence of agents so disposed can explain the absence of the convention of refraining from eating factory-farmed animals. If no agent (or only a few agents) see(s) factory-farmed animals as pitiable, such a convention is unlikely to be generated or sustained. The inverse may be true for Community *A*, i.e. there are sufficient agents in Community *A* that are disposed to feel compassion toward factory farmed animals. The presence of agents so disposed could have generated and sustained the convention of refraining from eating factory farmed animals. The presence/absence of factory-farmed animals as an eliciting condition of compassion can in turn be explained by the malleability of emotions. Emotions (including compassion) can be tuned. New eliciting conditions can be added, and existing conditions can be removed. As we can see, the bottom line here is that we can explain this instance of moral diversity with reference to emotion.

With the above in mind, let us now consider the conventionalist explanation of such an instance of moral diversity. For the conventionalist, the *presence/absence of certain conventions among communities can be explained with reference to agreements*. Before going any further, I want to clarify how the terms 'convention' and 'agreement' are used here. As Harman appears to use these terms interchangeably, I provide this clarification to avoid any confusion on their usage.⁷¹ Let us start with the term 'convention'. What I mean by convention thus far is 'a pattern of behaviour reproduced under the weight of precedent'. I am sticking to Millikan's definition of convention as outlined in chapter four. The focus here is on the pattern of behaviour involved. In our context,

⁷¹ See Harman 2000, 46 as an example.

the convention would be the *pattern of behaviour* of agents within Community *A* in refraining from eating factory farmed animals. Now let us turn our attention to the term 'agreement'. For Harman, we can understand an agreement as *shared intentions* of agents. Specifically, he argues that "there is an agreement in the relevant sense when each of a number of people has an intention on the assumption that others have the same intention" (Harman 2000, 26). In our context, the agreement may be an understanding among agents who are part of or share the views of Anonymous for the Voiceless. Each of these agents intend, *inter alia*, to refrain from eating factory farmed animals on the assumption that other agents also intend to do so. As we can see, an 'agreement' here refers to (a certain kind of) *shared intentions* rather than any pattern of behaviour. With this definitional framework outlined, we can proceed to examining how differing conventions (patterns of behaviour) can be explained with reference to agreements (shared intentions).

For the conventionalist, the existence of the convention of refraining from eating factory farmed animals in Community *A* can be explained by the existence of the shared intentions of (some) agents in Community *A* to refrain from such an act. It is the shared intentions of agents to act accordingly that generates and sustains the convention. Inversely, the absence of the same convention in Community *B* can be explained by the absence of any such shared intentions. Given that there are no shared intentions to refrain from eating factory farmed animals in Community *B*, such a dietary convention would not likely 'take off' in Community *B*. The differing shared intentions between Communities *A* and *B* can be explained by the differing interests that agents in these communities have. Agents in Community *A* may share an interest to protect factory farmed animals. They may want to further such an interest. One way of doing so is by seeking out like-minded agents. At this point, we see inklings of how an agreement (shared intentions) to refrain from eating factory farmed animals (and subsequently a convention) could have been formed in Community *A*. As for agents in Community *B*, one possibility is that they may not share such an interest to protect such animals. Hence, no such agreement or convention is likely to follow.

Conventionalism also does well in explaining moral diversity, especially in one respect, i.e. the theory accommodates a wide range of moral diversity. As argued earlier, for Harman, the presence/absence of certain conventions across/within communities can be explained with reference to the presence/absence of certain agreements across/within communities. If there are no

shared intentions by a group of agents to refrain from eating factory farmed animals in Community *B*, there would not be a convention where people refrain from eating such animals (and/or judgment that 'factory farming is wrong') in Community *B*. Environmental and/or societal factors can in turn explain the presence/absence of such agreements. For example, a community that suffers from frequent and severe famines may have too few agents who share interests in preventing animal cruelty for any agreement/convention to be in place/take off. The point I wish to emphasize here is that for Harman what unites the diverse conventions/judgments across/within communities are agreements. And given that Harman appears to give a 'free rein' on the content of such agreements, this allows conventionalism to account for a wide variety of moral prohibitions across and within communities including those on murder, incest, theft, abortion, gender reassignment surgeries, tax fraud, breast augmentations, pork consumption, beef consumption, etc. For conventionalism, the moral judgments that 'murder is wrong', 'breast augmentations are wrong', and 'eating pork is wrong' (varied though they may be) can have a common denominator, i.e. agreements. The conventions that criticize, punish and restrict such conduct share also the common feature of being grounded in agreements.

Conventionalism's ability to account for a wide variety of moral diversity appears to match the ability of EAMs to do the same. As Prinz argues, "[d]o immoral acts have anything in common? Ostensibly, the answer is no. Immoral acts comprise a hodgepodge: lying, stealing, hoarding, hurting, killing, neglecting, harassing, polluting, molesting, vandalizing, disrespecting, and so forth. *What do these things have in common other than the fact that we frown on all of them?*" (Prinz 2007, 48). Let us consider this argument in the context of chapters two, three and four. For EAMs, the abovementioned moral prohibitions (i.e. conventions that in one way or another criticize, discourage, punish, and restrict the abovementioned conduct) can be the 'product' of moral judgments that are grounded in anger, disgust, guilt, and/or shame. For example, in applying Zagzebski's account of moral judgment, the convention of ostracizing those who procure abortions (doctors and patients alike) could have arisen out of agents' seeing abortions as offensive/disgusting. And if we are to apply Prinz's account of moral judgment (as another example), the convention could have arisen out of agents' recognition that abortions are likely to trigger anger and/or guilt in them. As we can see, while EAMs account for a wide range of moral diversity with reference to emotions, conventionalism can do just as well with reference to agreements.

With the above section, we see a conventionalist explanation of moral unanimity and diversity. Conventionalism stands as a rival to the EAM model in at least two ways. The first point of rivalry is that, for the conventionalist, *agreements* (as opposed to emotion) are essentially involved in moral judgments. As argued in the preceding section, conventionalism offers an account of moral judgment that is indexed to agreements. The second point of rivalry is that the conventionalist model explains instances of moral unanimity and diversity with reference to *agreements* (as opposed to emotions). And at least in terms of explaining the variety of (certain kinds of) moral judgments/conventions, conventionalism appears to do just as well as EAMs. Now let us now consider the question *how does the conventionalist model lose out to the EAM model*.

Losing Out in Explaining Moral Unanimity

In this section, I argue that the conventionalist model loses out to the EAM model in at least the following way: The conventionalist explanation of moral unanimity is incomplete in the sense that (a) it needs to assume that communities across time and space came to the same kinds of agreements, and (b) this assumption stands in need of explanation and lacks an ‘agreement-based’ explanation. This incompleteness allows the EAM model to ‘steal the show’ in the sense that the model (i) offers a promising ‘emotion-based’ explanation on how agreements could have likely converged in the conventionalist model, and (ii) offers an ‘emotion-based’ explanation of moral unanimity without need of recourse to agreements.

Allow me to elaborate with reference to pancultural conventions on alleviating suffering. In chapter four, we saw that the communities of the ancient Israelites, Native Americans, and Australian aborigines shared conventions of alleviating suffering (in one form or another). For the conventionalist, we can explain this instance of moral unanimity with reference to the agreements that underlie these conventions shared by the Israelites, Native Americans, and Australian aborigines. The convergence of agreements in such communities is despite the fact that these communities span across different times, are in differing geographical locations, and were insulated from each other. The conventionalist must assume this feature in her explanation of (this instance of) moral unanimity. At this point, the conventionalist may step up and explain this

feature of convergence with reference to certain shared features of human life. More precisely, the conventionalist may say that our basic dispositions to promote survival and reciprocal relations, our recognition of our physical vulnerability and our odds of survival if we work together could have narrowed down the answers to the question *what should we do when other members of the community face a threat to their bodily integrity*. This, in turn, could have narrowed down the agreements that various communities could have arrived at.

While the above explanation may account for other instances of moral unanimity, it appears unable to fully account for the conventions of alleviating suffering shared by the Israelites, Native Americans, and Australian aborigines. As seen in chapter four, the Israelites appear to have left ‘forgotten sheaves’ for the sojourner. The Australian aborigines appear to have conventions that cared for the severely disabled. And the Native Americans took care of widows and elderly. A common feature of these conventions is that the sojourner, disabled, widow and elderly would not likely be able to ‘repay’ the kindness or would only be able to ‘repay’ in limited ways. In other words, the Israelites, Native Americans and Australian aborigines would appear to have conventions where they cared for such individuals regardless of whether the care given would be reciprocated. It would, as such, appear that the supposed narrowing down of the kinds of agreements that we can enter into (by ‘quid pro quo’ adjustments among members of communities) cannot adequately explain the convergence of agreements/conventions among the Israelites, Native Americans and Australian aborigines.

Perhaps the conventionalist may challenge the claim that the sojourner, disabled, widow and elderly cannot equally partake in the ‘quid pro quo’ arrangement. She may argue they may ‘give back’ in some other ways. For example, the elderly (being able to live up to a ripe old age) might have the wisdom of their experience to share. The physically disabled (who might have been a casualty of war) might have a different kind of wisdom to share. Caregivers as such would stand to ‘gain’ by alleviating the suffering of the elderly and disabled. If so, the narrowing down of agreements by ‘quid pro quo’ adjustments can still explain the convergence of agreements/conventions among the Israelites, Native Americans and Australian aborigines. This may be the case. But what if we slightly ‘tweak’ the situation at hand? What if the elderly and physically disabled that we are talking about are also mentally disabled and, as such, unable to

'give back' in the manner described above or in any other meaningful way? It would appear likely that the Israelites, Australian aborigines, and Native Americans would still have agreements/conventions to alleviate the suffering of such individuals. The feature of convergence in their agreements appears to still stand in need of a more adequate explanation.

At this juncture, the conventionalist would appear to have to abandon its 'completely conventional' or agreement-based explanation of (this instance of) moral unanimity. As seen above, Harman argues that we may be "genetically constructed" to feel concern for others (Harman 2000, 70). This claim may be inspired by Darwin's argument that it is the social instincts of an animal that leads it to feel sympathy for others and serve others in various ways (Darwin 1871: 2009, 72). And this conventionalist claim, as such, can be used to explain the above feature of convergence. Perhaps all healthy human beings have the disposition to feel compassion for those who are suffering. If so, then most Israelites, Australian aborigines and Native Americans would have the disposition to feel compassion for those who are suffering (even if they stand to gain little/nothing). Perhaps this shared disposition for compassion is what ultimately underlies our interest to alleviate the suffering of others. This, whether coupled with a recognition of our physical vulnerability or not, in turn, could have significantly narrowed down the kinds of agreements that the Israelites, Australian aborigines and Native Americans could have arrived at. This narrowing down of the kinds of agreements we enter into *by compassion* can thus explain the convergence of agreements on alleviating suffering in these communities despite the fact that they span across different times, are in differing geographical locations, and were insulated from each other. As we can see here, while the conventionalist can explain (instances of) moral unanimity, it does not appear to be able to explain the convergence of such agreements without appealing to an 'emotion-based' explanation.

The conventionalist may object that if different communities arrive at similar agreements, there must presumably be some empirical explanation for this. And given that conventionalism has provided an explanation of *what moral unanimity is* (i.e. different communities arriving at similar agreements), it need not provide an explanation of *how moral unanimity came about*. A response to this can be that there is still a need to explain how this convergence of agreements came about. Allow me to explain. First, the reader may note that it is plausible that (at least some) shared

conventions on alleviating suffering were not preceded by any such agreements. The shared motivational impulses (of members of a community) to alleviate suffering could have been sufficient to generate and sustain such conventions. I will address this with more detail later in this section. It would suffice to note at this point that while such shared motivational impulses can be understood as shared intentions on the part of the community's members to alleviate suffering, such shared intentions by themselves are not sufficient for there to be an agreement (as defined by conventionalism). For such an agreement to exist, there is still a need for the members of this community to assume that every other member intends to alleviate suffering. Second, if some shared conventions on alleviating suffering were not preceded by agreements, conventionalism has not yet provided an explanation of *what moral unanimity is* (in such cases). In other words, the conventionalist cannot say that moral unanimity is *the fact of different communities arriving at the same agreements* if some conventions on alleviating suffering were not preceded by any such agreements. Third, if conventionalism has not yet provided an explanation of what moral unanimity is (in cases where conventions were not preceded by agreements), it would still have to explain how moral unanimity came about (in such cases). And such an explanation would likely involve an 'emotion-based' explanation.

The above nicely segues us into a discussion on how conventionalism loses out to EAMs. While conventionalism would have to refer to emotion to explain (some) shared conventions, EAMs would not have to refer to agreements to explain such shared conventions. It is in this sense that conventionalism loses out. As I have shown how the conventionalist model needs to refer to emotion to explain the shared conventions, allow me now to briefly show how the EAM model need not refer to agreements to explain such conventions. As argued in chapter four, we can cluster the conventions on alleviating suffering shared by the Israelites, Australian aborigines and Native Americans around the shared emotion of compassion. In applying Zagzebski's account of moral judgment (as an example), if people across communities are disposed to see the common features in the frailty/helplessness of the elderly, the widow, the disabled, and/or the sojourner as coming under a thick affective concept of the pitiable, this can plausibly explain why the Israelites, Australian aborigines and Native Americans share conventions on alleviating suffering. This EAM explanation is 'more direct' when compared to the conventionalist explanation in the sense that the former 'skips' the need to introduce the concept of 'agreement' to explain such shared conventions. More precisely, the EAM model, in its explanation of (this instance of) moral

unanimity, need not introduce the proposition that the Israelites, Australian aborigines or Native Americans came to any agreement to alleviate suffering in the sense that each member of each of these communities intend to alleviate suffering on the understanding that others in each of these communities would do the same.

As it appears, the EAM theorist can explain moral unanimity without reference to agreements (whereas the conventionalist would not be able to do so). One question that can be raised at this juncture would be *why would this be an EAM advantage over conventionalism*. The answer can begin with the claim that EAM's explanation of moral unanimity is simpler than conventionalism's explanation. The conventionalist model appears to postulate two entities (emotion *and* agreement) in its hypothesis on moral unanimity. Whereas the EAM model postulates only one entity (emotion) in its hypothesis on moral unanimity. This parsimonious feature of the EAM model (when compared to the conventionalist model) would be an advantage (over the conventionalist model) in the sense that the EAM model would be more appealing to those who value theoretical simplicity, i.e. all other things being equal, the simpler explanation is to be preferred. And if "[m]ost philosophers believe that, other things being equal, simpler theories are better [emphasis added]" (Baker 2016), many would likely find the simpler EAM model more appealing (when compared to the conventionalist model). As it would appear, the simpler EAM model is to be preferred over conventionalist model in accordance with what E. C. Barnes argues to be the "anti-quantity principle", i.e. "if two competing theories tie on all other epistemic criteria, but one theory has fewer theoretical components than the other, the theory with fewer components is to be preferred" (Barnes 2000, 355).

Some find simplicity appealing on grounds that the simpler theory is more likely to be true or less likely to be false.⁷² Richard Swinburne, for example, argues that "other things being equal – the simplest hypothesis proposed as an explanation of phenomena is more likely to be the true one than is any other available hypothesis... is an ultimate *a priori* epistemic principle that simplicity is evidence for truth" (Swinburne 1997, 1). Here Swinburne's claim is that the *simplest* explanation

⁷² There are (obviously) other justifications for simplicity. To avoid doubt, I present this justification primarily to illustrate that there can be plausible justifications for simplicity. I do not intend to provide a complete defence of this justification nor do I intend to favour this justification over another.

of a phenomenon is to be preferred over other explanations of the same phenomenon as the former is more likely to be true. He does not justify this principle. Being the *ultimate* principle, the principle of simplicity cannot for Swinburne be justified by an even more basic principle. He supports his claim with an example (Swinburne 2001, 83). He argues first that formulae (a) $y=2x$ and (b) $y=2x+x(x-1)(x-2)(x-3)(x-4)(x-5)(x-6)z$ can explain the following observation on the nexus between x and y equally well:

x	0	1	2	3	4	5	6
y	0	2	4	6	8	10	12

Swinburne then goes on to argue that “[a]ll these different formulae although agreeing in yielding the values of y (for given values of x) observed so far, make totally different predictions for the future”. Despite this, we prefer formula (a) over (b) as (a) is the simpler one: “We believe [(a)] to be more likely to be true than any other formula of the stated form – as can be seen by the fact that we believe [(a)’s] predictions to be as such more likely to be true than those of any other formula of that form. If our life depended on predicting the correct value of y for $x = 9$, we would think it utterly irrational to make any prediction other than $y = 18$ ” (Swinburne 2001, 83-84). If it is correct that the *simpler* explanation of a phenomenon is to be preferred over another (more complex) explanation as the former is more likely to be true and if we apply them to our context, we can say that the simpler EAM model is to be preferred over the conventionalist model in their explanations of (an instance of) moral unanimity as the former is more likely to be true or less likely to be false (when compared to the latter).

At this point, one can still question *what makes the simpler EAM explanation more likely to be true or less likely to be false*. As Alan Baker argues, “the main problem with *a priori* justifications of simplicity is that it can be difficult to distinguish between an *a priori* defense and *no* defense(!)... It is unclear where leverage for persuading skeptics of the validity of such principles can come from, especially if the grounds provided are not themselves to beg further questions” (Baker 2016). In light of this, the claim that the simpler EAM explanation is more likely to be true or less likely to be false warrants at least a plausible answer to the question *what makes the simpler EAM explanation more likely to be true or less likely to be false*. An answer to this question can lie in the flip-side of a thesis in Karl Popper’s falsificationist philosophy. Adolf Grünbaum succinctly outlines the

Popperian thesis I have in mind as follows: “There are comparative *degrees of falsifiability among theories*, such that the more falsifiable of two given theories is the one having **greater content**. Indeed, the greater the content of a theory, the **bolder it is** by being more **risky and more likely to be false**” (Grünbaum 2008, 181). For Popper, theory *A* can have greater content than theory *B* to the extent that the former’s hypothesis is more ambitious in its explanation of the observable phenomena than that of the latter. For example, if the observable phenomena reveals that all *known* humans in the 1800s have died, hypothesis *A* (claiming that *all* humans would die) has ‘greater content’ than hypothesis *B* (claiming that all humans *born in the 1990s* would die). The former has greater content, is bolder, riskier, and *more likely to be false* (when compared to the latter). For Popper, hypothesis *A* is simpler than hypothesis *B* in the sense that the former “say[s] more about the world” (Fitzpatrick 2020) and, thereby, more likely to be false than the latter hypothesis that is more detailed and specific. Now there is a problem with such a view in that it appears to yield, at least in some instances, a counter-intuitive consequence. As Nelson Goodman argues, while the hypothesis “All maples are deciduous” has greater content than the hypothesis “All maples, except perhaps those in Eagleville, are deciduous”, the latter is more likely to be false (when compared to the former) (Goodman 1961, 150). One way to work around this apparent counter-intuitive consequence is to tweak what it means by ‘greater content’. The hypothesis that has ‘greater content’ can instead be taken to be a hypothesis that has more clauses than another hypothesis. In such a case, ‘all maples, except perhaps those in Eagleville, are deciduous’ would have greater content than “all maples are deciduous”. The former has greater content in the sense that it includes a further clause ‘except perhaps those in Eagleville’. By doing so, we can now say that the hypothesis with ‘greater content’ is more likely to be false (without inviting the above counter-intuitive consequence). When put this way, the latter hypothesis is simpler in the sense that it has lesser clauses and, thereby, less likely to be false when compared to the former hypothesis. This tweak would represent an obvious departure from Popper’s thesis. This is not a concern to me as it would serve my purposes in providing a reason why (at least in this instance) the simpler EAM model (the one with ‘lesser content’) can be preferred.

One may however be concerned that a simpler hypothesis (one which has a lesser content) need not be less likely to be false. For example, ‘all mammals give birth to live young’ is simpler (or has lesser content) than ‘all mammals, except monotremes, give birth to live young’. While the former has lesser content, it is the latter that offers a more accurate depiction of the situation. The point

here is that by having lesser content, we may leave out details crucial to the truth of a hypothesis. My response to this concern is that though it may apply to (many) other instances, this concern would not apply to the comparison between the EAM and conventionalist models. As shall be seen in the remaining parts of this chapter, the clause (that is part of the conventionalist model) excluded from the EAM model would appear to be unnecessary (and, thereby, would not be a detail that is crucial) in explaining the pancultural status of conventions on alleviating suffering.

Let us now consider this ‘tweak’ in the context of the EAM and conventionalist models. As argued earlier in this section, the conventionalist model appears to postulate two entities (emotion *and* agreement) in its hypothesis on moral unanimity. More precisely, the conventionalist model postulates that:

- (a) Humans across communities share a disposition to see the suffering of others as pitiable;
- (b) This shared disposition has led humans across communities to arrive at agreements on alleviating suffering; and
- (c) Claims (a) and (b) explain the conventions on alleviating suffering shared across communities.

As for the EAM model, it postulates one entity less than its conventionalist counterpart. More precisely, the EAM model postulates that:

- (i) Humans across communities share a disposition to see the suffering of others as pitiable;
- (ii) Claim (i) explains the conventions on alleviating suffering shared across communities.

With the above, we can see that the conventionalist model appears to be the *riskier* model with *greater content* (when compared to the EAM model). The conventionalist model postulates (a) and (b) to explain (an instance of) moral unanimity. While the EAM model need only postulate (i). The conventionalist model thus bears the risks of (b) being proven false. Whereas the EAM model bears no such additional risk. In light of this, the conventionalist model’s ‘greater content’ renders it more likely to be false (when compared to the EAM model). And the EAM model’s ‘lesser content’ (in the above respect) renders it more likely to be true (when compared to the conventionalist model). This can be a reason why the simpler EAM model is to be preferred.

At this point, one may object that while the EAM model does not *explicitly* introduce agreements into its explanation of (this instance of) moral unanimity, it does so *implicitly*. More precisely, the shared disposition to see the common features in the frailty/helplessness of the elderly, the widow, the disabled, and/or the sojourner as coming under a thick affective concept of the pitiable *by itself* would not be a complete explanation of why the Israelites, Australian aborigines and Native Americans have conventions on alleviating suffering. Agreements would appear to have to be implicitly introduced for the explanation to be complete. This appears to be especially so in light of Harman's argument that agreements are "normally arrived at *tacitly*, by mutual adjustment of different people's behaviour, *without conscious awareness* [emphasis added]" (Harman 2000, 68). If each member of each of these communities did not (tacitly) arrive at an agreement to alleviate suffering, each of these communities could not likely have had shared conventions on alleviating suffering. Even if everyone sees the elderly, the widow, the disabled, and/or the sojourner as pitiable, without the above (tacitly) shared intentions, such conventions on alleviating suffering could not likely have been sustained. In summary, the objection here is that the EAM model does not appear to be simpler (when compared to the conventionalist model). To explain moral unanimity, the EAM model would appear to have to also postulate (at least implicitly) that our shared disposition to see the suffering of others as pitiable has led humans across communities to arrive at agreements on alleviating suffering (before being able to explain the pancultural status of conventions on alleviating suffering).

One response to this can be to reject the proposition that agreements *as defined by conventionalism* are likely needed to generate or sustain conventions on alleviating suffering. Imagine a community comprising two agents, *A* and *B*. If *A* sees *B*'s suffering as pitiable (and vice versa), *A* would also likely have the motivational impulse to alleviate the suffering of *B* (and vice versa). This motivational impulse (if acted out by *A* and *B* often enough) could have been sufficient to generate and sustain the convention on alleviating suffering in this community of two. From a certain vantage point, it can be said that *A* and *B* have a *shared intention* to alleviate the suffering of each other (after all, they *both* have the motivational impulse to alleviate each other's suffering). But this shared intention is not a sufficient condition for an agreement as defined by conventionalism to exist. There is still another condition to be satisfied. For there to be an agreement, the shared intention of *A* and *B* to alleviate each other's suffering must be coupled with the understanding that *A/B* intends to alleviate *B/A*'s suffering. This understanding on *A*'s and *B*'s part, however,

does not appear to be needed for a convention on alleviating suffering to be generated or sustained in this community of two. This hypothetical example can also be extended to communities consisting of greater numbers of agents.⁷³

At this juncture, Mencius' argument in 2A6 would be worth reiterating: "[a]ll human beings have a mind that cannot bear to see the suffering of others... Now, if anyone were suddenly to see a child about to fall into a well, his mind would be filled with... compassion. That he would react accordingly is not because he would hope to use the opportunity to ingratiate himself with the child's parents, nor because he would seek commendation from neighbors and friends, nor because he would hate the adverse reputation [that could come from not reacting accordingly]" (Mencius c 300 BC: 2009, 2A6). As we can see here, an agent's compassion directed at the (anticipated) suffering of the child is sufficient to motivate her to save the child. The motivational impulse to save the child can be there without any thought of gain on the agent's part. We can extend this claim further to say that the agent can have the motivational impulse to save the child without any understanding/expectation (on the agent's part) that the child or her parents would intend to do the same if the agent herself was (about to be) in distress. And if sufficient agents feel compassion toward the suffering of others and, by extension, are motivated to alleviate their suffering (without any thought, understanding or expectation that others would intend to reciprocate), there appears to be no reason why a convention on alleviating suffering cannot be thus generated and sustained.

⁷³ The conventionalist may at this point insist that some kind of mutual recognition is likely to be essential to sustain a convention in larger communities. More precisely, she may insist that it would seem that conventions of larger communities would have to be sustained not only by shared intentions of its members to alleviate suffering, but also by the understanding (on the part of its members) that every other member intends to alleviate suffering. While it may work for a community of two or ten to sustain a convention on alleviating suffering without such an understanding, it would appear implausible for a community of 50, 500, 5000, etc. to sustain such a convention without such an understanding. One response to this is that there are means of sustaining such conventions in larger communities other than that of the understanding (on the part of their members) that every other member intends to alleviate suffering. Deviations (by members) from the expectations that come with the convention on alleviating suffering can (for example) be met with punishments. As long as a sizeable number of members within that larger community have shared intentions to alleviate suffering, this coupled with a collective effort (of community members) to punish deviants would likely be sufficient to sustain the convention on alleviating suffering. Some community members could also have developed an unthinking habit of alleviating the suffering of others. It would seem that this habit could also help sustain the convention without the need for any mutual recognition (even when they no longer feel the strong motivational impulse to alleviate suffering).

Allow me to proffer more details by drawing the reader's attention (again) to the meaning of 'convention' adopted in chapter four, i.e. patterns of behaviour reproduced under the weight of precedent (Millikan 2005). With this in mind, we can envision a community where:

- (i) A number of agents within a community feel compassion toward the suffering of others;
- (ii) These agents are motivated via compassion to alleviate the suffering of others; and
- (iii) These agents (act on their motivation to) alleviate the suffering of others and continue to alleviate suffering on the weight of their own precedent.

As we can see here, the convention of alleviating suffering (at (iii) above) can be generated in a community without its agents having any thought, understanding or expectation that other agents would intend to reciprocate. The convention (at (iii) above) can also be sustained by (i) and (ii). As long as there are agents that feel compassion and are motivated to alleviate suffering, it is likely that an existing convention on alleviating suffering within a community would persist. There are two further points I wish to make. First, as we can see with (i) to (iii), agreements as defined by conventionalism need not be postulated (even implicitly) by the EAM model in its explanation of moral unanimity. Secondly, the conventionalist model would appear to need to introduce (i), (ii) and (iii) in its explanation of (such an instance of) moral unanimity. Allow me to explain with the following:

- (A) A number of agents within a community feel compassion toward the suffering of others;
- (B) These agents are motivated via compassion to alleviate the suffering of others;
- (C) These agents, as a result of (A) and (B), arrive at an agreement to alleviate suffering; and
- (D) These agents, as a result of (C), (act on their motivation to) alleviate the suffering of others and continue to alleviate suffering on the weight of their own precedent.

As we can see, (A), (B) and (D) (almost) mirror (i), (ii) and (iii). (A) and (B) also do not appear to be able to be discarded from the conventionalist explanation of moral unanimity. And while (C) is obviously needed in the conventionalist explanation, the EAM model is free to discard it. If this analysis is correct, the EAM model remains as the simpler model (when compared to the conventionalist model).

However, one may still insist that the EAM model is not free to discard agreements from its explanation of moral unanimity. While the EAM model can explain the formation of shared conventions without agreements, agreements appear to be a salient feature of human social life. As discussed in chapter four, Deuteronomy 24:19 instructs the ancient Israelites to leave their forgotten sheaves for strangers, orphans and widows. We can construe this instruction as a written manifestation of an agreement. More precisely, it can be construed as a codification of the Israelites' intention to care for strangers, their orphans and widows on the assumption that other Israelites intend to do the same. Now allow me to consider a contemporary example. Caritas Internationalis is a confederation of Catholic charitable organisations across the world. Its code of ethics states, inter alia, that: "... united in one human family we are profoundly moved by the suffering of others and have a moral duty to recognise the humanitarian imperative to respond. This duty is essential both to our identity as a Catholic organization and to our membership of the human family. Thus, as members of the international community, we recognise our obligation to provide humanitarian assistance..." (Caritas Internationalis 2014). What I wish to tease out of this statement is that it can be understood as a written manifestation of an agreement: where members of Caritas express their intention to provide humanitarian aid where needed on the assumption that other members intend to do the same. As we can see, these are two clear examples of how agreements may be involved in our shared conventions on alleviating suffering. They are 'clear' in the sense that the apparent agreements are made manifest in writing. I would think there may be many other instances of agreements where no writing is involved. The Australian aborigines Dalebura tribe, for example, had the convention of caring for their disabled. The Native Americans too had the convention of caring for their widows, orphans and elderly. While such conventions were not codified, it may be the case that in such instances, the Australian aborigines and Native Americans had agreements in the sense that they intended to care for those who were vulnerable on the assumption that other members of their communities intend to do the same.

What is to be emphasised at this point is that agreements are likely to be involved in many instances of shared conventions. Agreements can be involved in the sense that they are the *result* of solely emotion-based conventions. And more precisely, agreements can follow from the process set out in (i) to (iii) above. Allow me to explain this in detail by using/borrowing Wilfrid Sellars'

Myth of Jones as a starting point. Sellars' myth begins with Rylean ancestors and their language that is limited to a "descriptive vocabulary [that] speaks of *public* properties of *public* objects ... [emphasis added]". However, at this point, these ancestors do not have the vocabulary "to speak of *inner* episodes and immediate experiences [emphasis added]" (Sellars 1991, 178, 179). For example, our Rylean ancestors could have uttered 'that fish is big', but not '*it looks to us* that that fish is big'. The latter involves an 'inner' vocabulary, while the former does not. Sellars then proceeds to show how our Rylean ancestors could have started with a vocabulary limited to 'public' objects and ended up with a more expansive vocabulary that encompasses their 'inner episodes'. He does so by placing Jones, a genius, among our Rylean ancestors. He observed that our ancestors continued to "behave intelligently not only when their conduct is threaded on a string of overt verbal episode – that is to say ... when they 'think out loud' – but also when no detectable verbal output is present" (Sellars 1991, 186). It was this that led Jones to conclude that they must be experiencing 'inner episodes' (i.e. "[s]eeing that something is the case") allowing them to behave intelligently even when they are silent (Sellars 1991, 190). Jones then communicates this to our Rylean ancestors and trains them to say things like 'it looks to us that that fish is big', allowing them to realise their self-awareness and ability to report their 'inner episodes'.

The points I wish to tease out from this myth are that: (a) the 'public' vocabulary of our Rylean ancestors was sufficient for Rylean communication; (b) Jones' subsequent introduction of an 'inner' vocabulary was (obviously) not essential to early Rylean communication; (c) despite this non-essentiality, the 'inner' vocabulary greatly enhanced Rylean communication and becomes 'part and parcel' of it; and (d) this enhanced Rylean communication would likely have been passed down through the generations. And in light of (b) to (d), it would, thereby, not make sense to discard this 'inner' vocabulary from an explanation of Rylean communication. Let us now apply these points to the EAM explanation of moral unanimity. For the EAM model, the compassion of our EAM ancestors was sufficient to allow them to (act on their motivation to) alleviate the suffering of others and continue to alleviate suffering on the weight of their own precedent. And the agreements (that come after the development of such conventions on alleviating suffering) are not essential to the development of such conventions. This being said, such agreements would likely have enhanced the viability of such conventions. For example, if the Israelites (apart from having the motivation to alleviate the suffering of strangers, orphans, and widows) had the shared intentions to alleviate their suffering, these shared intentions would have likely helped sustain

their convention on alleviating suffering. Such shared intentions would have subsequently become 'part and parcel' of the convention on alleviating suffering. In this instance, these shared intentions can be said to have been codified in the *Book of Deuteronomy* and passed down through the generations. The problem of the EAM model now becomes evident. Given that subsequent generations would have inherited a convention enhanced by agreement, it would not make sense for the EAM model to discard agreements from its explanation of this and other such instances of shared conventions. In other words, the EAM model would still need agreements to explain such instances of shared conventions. If so, we cannot say that the EAM model is simpler when compared to the conventionalist model.

An EAM response to this problem can start by accepting that (at least in many instances) agreements can follow from the process set out in (i) to (iii) above, enhance the viability of our shared conventions, and, become 'part and parcel' of our shared conventions. The EAM theorist can also accept that such agreement-enhanced conventions could have been inherited by subsequent generations and agreements would still be a much needed 'entity' to explain such conventions. The EAM theorist can accept all these while holding on to the position that the EAM model is still simpler (when compared to the conventionalist model). More precisely, the EAM model remains simpler in the sense that it does not posit more entities than what is necessary to explain *each type* of moral unanimity.

Allow me to explain. The objection above made salient the existence of shared conventions that are agreement-enhanced. And in such instances, the EAM model (like the conventionalist model) would have to posit the entity of 'agreements' to explain them. As such, the EAM model cannot be said to be simpler than the conventionalist model in explaining agreement-enhanced conventions. This being said, there may be other instances of shared conventions that are not agreement-enhanced. I start by borrowing an argument from Hume:

"A tacit promise is, where the will is signified by other more diffuse signs than those of speech; but a will there must certainly be in the case, and that can never escape a person's notice, who exerted it, however silent or tacit. But were you to ask the far greatest part of the nation, whether they had ever consented to the authority of their rulers, or promis'd to

obey them, they wou'd be inclin'd to think very strangely of you; and wou'd certainly reply, that the affair depended not on their consent, but that they were born into such obedience" (Hume 1896, 547-548)

Hume's argument above is directed against those who argue that agreements could have been arrived at implicitly. For him, even an implicit agreement involves wilful consent on the part of its parties. He then argues that most who are regarded as bound by a political agreement did not give their wilful consent; most were not given the choice to decide whether to opt in. Now this argument can be co-opted to show that there may be shared conventions that are not agreement-enhanced. We can start by thinking in terms of Harman's *moral* agreements, instead of Hume's *political* agreements. We can then proceed by replacing Hume's 'wilful consent' with Harman's 'intention to act in certain ways on the assumption that others intend the same'. Once this is done, the Humean argument can be understood as follows: Even though moral agreements can be implicitly entered into, each party to such agreements must intend to act in x ways on the assumption that other parties intend to do the same. However, it would seem implausible that all who act in x ways would have such intentions. Many who act in x ways may do so out of a sheer habit (for example). The point here being that while agreements may be a salient feature of *many* shared conventions, agreements need not be a part of every *shared* convention. As such, there is likely to be some shared conventions that are not enhanced by agreement. If so, the EAM model would not have to posit the entity of 'agreements' to explain them. In this light, the EAM model can still be said to be simpler than the conventionalist model (in the sense that the former postulates one less entity than the latter) in explaining shared conventions that are *not agreement-enhanced*.

Final Thoughts

With the above, we see conventionalism as a rival to EAMs in at least two ways. First, conventionalism puts forward what appears to be a plausible account of moral judgment. For EAMs, *emotion* is (in one way or another) essentially involved in moral judgments. And for conventionalism, *agreements* are essentially involved in moral judgments. Second, conventionalism proffers an explanation of instances of moral unanimity and diversity. For EAMs, moral unanimity and diversity can be explained with reference ultimately to (some) emotions. As for conventionalism, they can be explained ultimately with reference to agreements. This comparative

exercise is a novelty of this chapter as I found no such comparison between EAMs and conventionalism on their explanations of moral unanimity and diversity in the existing philosophical literature. Another novelty is my argument that the EAM model had at least one advantage over the conventionalist model, i.e. while conventionalism would have to refer to emotion to explain instances of moral unanimity, EAMs would not have to refer to agreements to explain such instances. This is an advantage over the conventionalist model as the EAM model with its apparent simplicity would, other things being equal, likely attract those who see simplicity as a theoretical virtue. With this in mind, I hope that the reader is now inclined to accept that not only can the EAM model explain moral unanimity and diversity (and this counts as a reason in favour of this model), the model can also best a rival conventionalist model in at least one respect.

Bibliography

- Appiah, Kwame Anthony. 2008. *Experiments in Ethics*. Cambridge: Harvard University Press.
- Austin, John. 1995: 2001. *The Province of Jurisprudence Determined*. Edited by Wilfrid E. Rumble. New York: Cambridge University Press.
- Baker, Alan. 2016. "Simplicity." Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy (Winter 2016 Edition)*.
- Barnes, E. C. 2000. "Ockham's Razor and the Anti-Superfluity Principle." *Erkenntnis* 53 (3): 353-374.
- Barsam, Richard Meran. 1973: 1992. *Non-fiction Film: A Critical History*. Bloomington: Indiana University Press.
- Ben-Ze'ev, Aaron. 2010. "The Thing Called Emotion." In *The Oxford Handbook of Philosophy of Emotion*, edited by Peter Goldie, 41-62. New York: Oxford University Press.
- Blackburn, Simon. 2006. "Must We Weep for Sentimentalism ." In *Contemporary Debates in Moral Theory*, by James Dreier, 145-159. Oxford: Blackwell Publishing .
- . 2002. "Précis of Ruling Passions." *Philosophy and Phenomenological Research* LXV (1): 122-135.
- . 1998. *Ruling Passions: A Theory of Practical Reasoning*. New York: Oxford University Press.
- Brady, Michael S. 2013. *Emotional Insight: The Epistemic Role of Emotional Experience*. New York: Oxford University Press.
- Brentano, Franz. 1874: 1995. *Psychology from an Empirical Standpoint*. Edited by Linda L. McAlister. Translated by Antos C. Rancurello, D.B. Terrell and Linda L. McAlister. New York: Routledge.
- Burke, Edmund. 1909-14. *Reflections on the French Revolution*. Vols. XXIV, Part 3. New York: P.F. Collier & Son.
- Bykvist, Krister. 2009. "No Good Fit: Why the Fitting Attitude Analysis of Value Fails." *Mind* 118 (469): 1-30.
- Calhoun, Cheshire. 2001. "Making Up Emotional People: The Case for Romantic Love." In *The Passions of Law*, edited by Susan Bandes, 217-240. New York: NYU Press.
- Campbell, Joseph. 1949: 2004. *The Hero With A Thousand Faces*. Princeton: Princeton University Press.
- Cannon, Walter. 1929. *Bodily Changes in Pain, Hunger, Fear and Rage*. 2nd Edition. New York: Appleton.
- Caritas Internationalis. 2014. *Code of Ethics for Caritas Internationalis member organisations*. 18 May. Accessed May 9, 2020. <https://www.caritas.eu/code-of-ethics/>.

- Cleckley, Hervey. 1941. *Mask of Sanity: An Attempt to Reinterpret the So-called Psychopathic Personality*. St Louis: C.V. Mosby Company.
- Confucius. c 500 BC: 2003. *Analects*. Translated by Edward Slingerland. Indianapolis: Hackett Publishing Company.
- Copp, David. 1982. "Harman on Internalism, Relativism and Logical Form." *Ethics* 227-242.
- Cosmides, Leda, Ricardo Andrés Guzmán, and John Tooby. 2019. "The Evolution of Moral Cognition." In *The Routledge Handbook of Moral Epistemology*, 174-228. New York: Routledge.
- Cullity, Garrett. 2011. "Moral Judgement." *Routledge Encyclopedia of Philosophy* (Taylor and Francis).
- Döring, Sabine A. 2014. "Why Recalcitrant Emotions Are Not Irrational." In *Emotion and Value*, edited by Sabine Roeser and Cain Todd, 124-136. New York: Oxford University Press.
- Damasio, Antonio. 1995. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: First Avon .
- D'Arms, Justin, and Daniel Jacobson. 2006. *Anthropocentric Constraints on Human Value*. Vol. 1, in *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 99-126. New York: Oxford University Press.
- Darwall, Stephen, Allan Gibbard, and Peter Railton. 1992. "Toward Fin de siècle Ethics: Some Trends." *The Philosophical Review* 101 (1): 115-189.
- Darwin, Charles. 1872: 1965. *The Expression of the Emotions in Man and Animals*. Chicago: University of Chicago Press.
- . 1871: 2009. *The Descent of Man and Selection in Relation to Sex*. New York: Cambridge University Press.
- de Tocqueville, Alexis. 1835: 2010. *Democracy in America*. Edited by Eduardo Nolla. Translated by James T. Schleifer. Vol. 1. Indianapolis: Liberty Fund, Inc.
- de Waal, Frans. 2014. "Natural Normativity: The "Is" and "Ought" of Animal Behavior." In *Evolved Morality: The Biology and Philosophy of Human Conscience* , 49-68. Leiden: Brill.
- . 2006. *Primates and Philosophers*. Edited by Stephen Macedo and Josiah Ober. Princeton: Princeton University Press.
- de Waal, Frans, and A. van Roosmalen. 1979. "Reconciliation and Consolation Among Chimpanzees." *Behavioral Ecology & Sociobiology* 5: 55-66.
- Deigh, John. 2009. "Concepts of Emotions in Modern Philosophy and Psychology." In *The Oxford Handbook of Philosophy of Emotion*, edited by Peter Goldie. New York: Oxford University Press.
- . 2008. *Emotions, Values, and the Law*. New York: Oxford University Press.
- Deonna, Julien, Christine Tappolet, and Fabrice Teroni. 2015. "Emotion, Philosophical Issues About." *WIREs Cognitive Science* (6): 193-207.
- Dewey, John. 1895. "The Theory of Emotion." *The Psychological Review* 2 (1): 13-32.

- . 1894. "The Theory of Emotion." *The Psychological Review* 1 (6): 553-569.
- Dimberg, Ulf, Monika Thunberg, and Kurt Elmehed. 2000. "Unconscious Facial Reactions to Emotional Facial Expressions." *Psychological Science* 11 (1): 86-89.
- Edmonds, David. 2014. *Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us about Right and Wrong*. Princeton: Princeton University Press.
- Eisenberg, Nancy. 2000. "Emotion, Regulation, and Moral Development." *Annual Review of Psychology* 51: 665-697.
- Ekman, Paul. 1980. "Biological and Cultural Contributions to Body and Facial Movement in the Expression of Emotions." In *Explaining Emotions*, 73-102. Los Angeles: University of California Press.
- Ekman, Paul, and Daniel Cordaro. 2011. "What is Meant by Calling Emotions Basic." *Emotion Review* 3 (4): 364-370.
- Elster, Jon. 1999. *Alchemies of the Mind*. Cambridge: Cambridge University Press.
- Epictetus. c 108 AD: 1904. *Discourses of Epictetus*. Translated by George Long. New York: D. Appleton and Company.
- Feinberg, Matthew, Olga Antonenko, Robb Willer, E. J. Horberg, and Oliver P. John. 2014. "Gut Check: Reappraisal of Disgust Helps Explain Liberal–Conservative Differences on Issues of Purity." *Emotion* 14 (3): 513-521.
- Finnis, John. 2011. *Natural Law and Natural Rights*. New York: Oxford University Press.
- Fisk, Robert. 2010. *The Independent*. 10 September. Accessed January 2, 2020. <https://www.independent.co.uk/voices/commentators/fisk/robert-fisk-the-truth-about-honour-killings-2075317.html>.
- Fitzpatrick, Simon. 2020. "Simplicity in the Philosophy of Science." *Internet Encyclopedia of Philosophy*. Accessed April 22, 2020. <https://www.iep.utm.edu/simplici/#H6>.
- Flanagan, Owen. 2016. *The Geography of Morals: Varieties of Moral Possibility*. New York: Oxford University Press.
- . 2014. *Moral Sprouts and Natural Teleologies: 21st Century Moral Psychology Meets Classical Chinese Philosophy*. Milwaukee: Marquette University Press.
- Frankfurt, Harry G. 2006. *Taking Ourselves Seriously & Getting it Right*. Stanford: Stanford University Press.
- Gert, Joshua. 2015. "Moral Sentimentalism." *Routledge Encyclopedia of Philosophy*. doi:doi:10.4324/9780415249126-L3578-1.
- Gibbard, Allan. 1993. "Précis of Wise Choices, Apt Feelings." *Philosophical Issues* 4 (Naturalism and Normativity): 33-35.

- . 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Oxford University Press.
- Goetz, Jennifer L., Dacher Keltner, and Emiliana Simon-Thomas. 2010. "Compassion: An Evolutionary Analysis and Empirical Review." *Psychological Bulletin* 136 (3): 351-374.
- Goldie, Peter. 2002. "Emotions, Feelings and Intentionality." *Phenomenology and the Cognitive Sciences* 1 (3): 235-254.
- . 2000. *The Emotions: A Philosophical Exploration*. New York: Oxford University Press.
- Goodman, Nelson. 1961. "Safety, Strength, Simplicity." *Philosophy of Science* 150-151.
- Grünbaum, Adolf. 2008. "Is Simplicity Evidence of Truth?" *American Philosophical Quarterly* 45 (2): 179-189.
- Greenspan, Patricia. 1988. *Emotions and Reasons: An Inquiry into Emotional Justification*. New York: Routledge.
- Gross, James J., and Oliver P. John. 2003. "Individual differences in two emotion regulation processes: Implications for affect, relationships, and wellbeing." *Journal of Personality and Social Psychology* 85 (2): 348-362.
- Grzankowski, Alex. 2016. "Attitudes Towards Objects." *Nous* 50 (2): 314-328.
- . 2016. "The Real Trouble with Recalcitrant Emotions." *Erkenntnis* 82: 641-651.
- Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People Are Divided By Politics and Religion*. London: Penguin Books.
- . 2001. "The Emotional Dog and Its Rational Tail." *Psychological Review* 108: 814-834.
- Haidt, Jonathan, Fredrik Bjorklund, and Scott Murphy. 2000. "Moral Dumbfounding: When Intuition Finds No Reason." *Unpublished Manuscript*.
- Hamlin, J. Kiley, Karen Wynn, and Paul Bloom. 2007. "Social Evaluation by Preverbal Infants." *Nature* 557-560.
- Harman, Gilbert. 2000. *Explaining Value and Other Essays in Moral Philosophy*. Oxford: Oxford University Press.
- . 1996. "Social Contracts." In *Moral Relativism and Moral Objectivity*, by Gilbert Harman and Judith Jarvis Thomson, 20-31. Cambridge, MA: Blackwell Publishers.
- Harner, Michael. 1977. "The Ecological Basis for Aztec Sacrifice." *American Ethnologist* 4 (1): 117-135.
- Hatfield, Elaine, John T. Cacioppo, and Richard L. Rapson. 1993. "Emotional Contagion." *Current Directions in Psychological Science* 2 (3): 96-99.
- Heitman, Elizabeth, and Sergio Litewka. 2011. "International perspectives on plagiarism and considerations for teaching international trainees ." *Urologic Oncology: Seminars and Original Investigations* (29): 104-108.

- Helm, Bennett W. 2009. "Emotions as Evaluative Feelings." *Emotion Review* 248-255.
- . 2001. *Emotional Reason: Deliberation, Motivation and the Nature of Value*. New York: Cambridge University Press.
- Herd, Gilbert H. 1987. *The Sambia: Ritual and Gender in New Guinea*. New York: Holt, Rinehart, and Winston.
- Herek, Gregory M. 1993. "Documenting prejudice against lesbians and gay men on campus: the Yale Sexual Orientation Survey." *Journal of Homosexuality* 25 (4): 15-30.
- Hobbes, Thomas. 1651: 1965. *Leviathan*. Oxford: Clarendon Press.
- Hochschild, Arlie Russell. 1983. *The Managed Heart: Commercialization of Human Feeling*. Berkeley and Los Angeles: University of California Press.
- Hume, David. 1896. *Treatise of Human Nature*. Edited by L. A. Selby-Bigge. Oxford: Clarendon Press.
- . 1777:1975. *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Oxford: Oxford University Press.
- . 1777: 1912. *An Enquiry Concerning the Principles of Morals*. Project Gutenberg. Accessed November 2019. https://www.gutenberg.org/files/4320/4320-h/4320-h.htm#2H_4_0003.
- Ihara, Craig K. 1991. "David Wong on Emotions in Mencius." *Philosophy East and West* 41 (1): 45-53.
- ILGA, and Aengus Carroll. 2016. *State Sponsored Homophobia*. Geneva: ILGA.
- Ivanhoe, Phillip J. 1990. *Ethics in the Confucian Tradition*. Atlanta: Scholar's Press.
- Jacobson, Daniel. 2011. "Fitting Attitude Theories of Value." Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*. <<https://plato.stanford.edu/archives/spr2011/entries/fitting-attitude-theories/>>.
- James, Scott M. 2011. *An Introduction to Evolutionary Ethics*. Oxford: Wiley-Blackwell.
- James, William. 1884. "What is an Emotion?" *Mind* (9): 188-205.
- Johnston, Mark. 2001. "The Authority of Affect." *Philosophy and Phenomenological Research* 63 (1): 181-214.
- Kahneman, Daniel. 2011: 2012. *Thinking, Fast and Slow*. Penguin Books.
- Kass, Leon R. 1998. "The Wisdom of Repugnance: Why We Should Ban the Cloning of Humans." *Valparaiso University Law Review* 32 (2): 679-705.
- Kaufman, Whitley. 2007. "Karma, Rebirth, and the Problem of Evil: A Reply to Critics." *Philosophy East and West* 57 (4): 556-560.
- Kauppinen, Antti. 2016. "Moral Sentimentalism." *The Stanford Encyclopedia of Philosophy*.
- Kerstein, Samuel J. 2006. "Reason, Sentiment, and Categorical Imperatives." In *Contemporary Debates in Moral Theory*, edited by James Dreier, 129-143. Oxford: Blackwell Publishing.

- Kim, Myeong-seok. 2018. "Emotion and Judgment: Two Sources of Moral Motivation in Mèngzǐ." *Dao* 17: 51-80.
- . 2014. "Is There No Distinction Between Reason and Emotion in Mengzi." *Philosophy East and West* 64 (1): 49-81.
- Kitcher, Philip. 2014. "Is a Naturalized Ethics Possible?" *Behaviour* 151: 245-260.
- Korsgaard, Christine M. 2010. "Reflections on the Evolution of Morality." *The Amherst Lecture in Philosophy* 5 1-29.
- . 1996: 2010. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental evidence of massive-scale emotional contagion through social networks." *Proceedings of the National Academy of Sciences of the United States of America* 111 (24): 8788-8790.
- Lewis, Clive Staples. 1944: 2001. *The Abolition of Man*. New York: Harper Collins.
- Lewis, David. 1969: 2002. *Convention: A Philosophical Study*. Oxford: Blackwell Publishers.
- MacIntyre, Alasdair. 2013. "On Having Survived the Academic Moral Philosophy of the Twentieth Century? Philosophical Essays in Honor of Alasdair MacIntyre." In *What Happened in and to Moral Philosophy in the Twentieth Century?*, edited by Fran O'Rourke, 17-34. Notre Dame, IN: University of Notre Dame Press.
- Manjoo, Rashida. 2012. *Report of the Special Rapporteur on Violence Against Women, its Causes and Consequences*. Human Rights Council, United Nations.
- Marcus, Gary. 2004. *The Birth of the Mind: How a Tiny Number of Genes Creates the Complexities of Human Thought*. New York: Basic Books.
- McRae, Emily. 2011. "The Cultivation of Moral Feelings and Mengzi's Method of Extension." *Philosophy East and West* 61 (4): 587-608.
- Mencius. c 300 BC: 2008. *Mengzi*. Translated by Bryan W. Van Norden. Indianapolis: Hackett Publishing Company.
- . c 300 BC: 2009. *Mencius*. Translated by Irene Bloom. New York: Columbia University Press.
- Miller, William Ian. 1997. *The Anatomy of Disgust*. Cambridge, MA: Harvard University Press.
- Millikan, Ruth Garrett. 2005. *Language: A Biological Model*. New York: Oxford University Press.
- Molewijk, Bert, Dick Kleinlugtenbelt, and Guy Widdershoven. 2011. "The Role of Emotions in Moral Case Deliberation: Theory, Practice, and Methodology." *Bioethics* 25 (7): 383-393.
- Morrison, Todd G., Mark J. Kiss, CJ Bishop, and Melanie A. Morrison. 2019. "'We're Disgusted With Queers, not Fearful of Them': The Interrelationships Among Disgust, Gay Men's Sexual Behavior, and Homonegativity." *Journal of Homosexuality* 66 (7): 1014-1033.
- Morton, Adam. 2009. "Epistemic Emotions." In *The Oxford Handbook of Philosophy of Emotion*, edited by Peter Goldie, 385-400. New York: Oxford University Press.

- Neal, David T., and Tanya L. Chartrand. 2011. "Embodied Emotion Perception: Amplifying and Dampening Facial Feedback Modulates Emotion Perception Accuracy." *Social Psychological and Personality Science* 2 (6): 673-678.
- Neu, Jerome. 2004. "Emotions, Action and Freedom." In *Thinking About Feeling: Contemporary Philosophers on Emotions*, edited by Robert C. Solomon, 275-305. New York: Oxford University Press.
- . 2000. *A Tear is an Intellectual Thing: The Meaning of Emotions*. New York: Oxford University Press.
- Nichols, Shaun. 2008. *Sentimentalism Naturalized*. Vol. 2, in *Moral Psychology: The Cognitive Science of Morality: Intuition and Diversity*, edited by Walter Sinnott-Armstrong. Cambridge: MIT Press.
- Nivison, David S. 1996. *The Ways of Confucianism: Investigations in Chinese Philosophy*. Edited by Bryan W. Van Norden. Chicago and La Salle: Open Court.
- . 1980. "Two Roots or One?" *Proceedings and Addresses of the American Philosophical Association* 53 (6): 739-761.
- Nussbaum, Martha C. 2016. *Anger and Forgiveness: Resentment, Generosity, Justice*. New York: Oxford University Press.
- . 2004. *Hiding from Humanity: Disgust, Shame, and the Law*. Princeton: Princeton University Press.
- . 2001. *Upheavals of Thought: The Intelligence of Emotions*. New York: Cambridge University Press.
- . 1988. "Non-Relative Virtues: An Aristotelian Approach." *Midwest Studies in Philosophy* XIII: 32-53.
- Ochsner, Kevin N., and James J. Gross. 2008. "Cognitive Emotion Regulation: Insights From Social Cognitive and Affective Neuroscience." *Current Directions in Psychological Science* 17 (2): 153-158.
- Oddie, Graham. 1999. "Moral Realism, Moral Relativism and Moral Rules (A Compatibility Argument)." *Synthese* 117: 251-274.
- Olsen, Vegard V, Ricardo G Lugo, and Stefan Sütterlin. 2015. "The Somatic Marker Theory in the Context of Addiction: Contributions to Understanding Development and Maintenance." *Psychology Research and Behavior Management* 8: 187-200.
- Pascal, Blaise. 1670: 1995. *Pensees and Other Writings*. Translated by Anthony Levi. New York: Oxford University Press.
- Perkins, Franklin. 2002. "Mencius, Emotion, and Autonomy." *Journal of Chinese Philosophy* 207-226.
- Pettigrove, Glen, and Koji Tanaka. 2014. "Anger and Moral Judgment." *Australasian Journal of Philosophy* 92 (2): 269-286
- Pettigrove, Glen, and Nigel Parsons. 2012. "Shame: A Case Study of Collective Emotion." *Social Theory and Practice* 38 (3): 504-530.
- Piaget, Jean. 1932. *The Moral Judgment of the Child*. Glencoe, IL: The Free Press.

- Prinz, Jesse. 2008. "Precis of Gut Reactions." *Philosophy and Phenomenological Research* LXXVI (2008): 707-711.
- . 2007. *The Emotional Construction of Morals*. New York: Oxford University Press.
- . 2006. "The Emotional Basis of Moral Judgments." *Philosophical Explorations* 29-43.
- . 2004. *Gut Reactions: A Perceptual Theory of Emotion*. New York: Oxford University Press.
- Rawls, John. 1951. "Outline of a Decision Procedure for Ethics." *The Philosophical Review* 60 (2): 177-197.
- Reimann, Martin, and Antoine Bechara. 2010. "The Somatic Marker Framework as a Neurological Theory of Decision-making: Review, Conceptual Comparisons, and Future Neuroeconomics Research." *Journal of Economic Psychology* 31: 767-776.
- Roberts, Robert. 2013. *Emotions in the Moral Life*. New York: Cambridge University Press.
- . 2009. "Emotions and the Canons of Evaluation." In *The Oxford Handbook of Philosophy of Emotion*, edited by Peter Goldie, 561 - 583. Oxford University Press.
- . 2003. *Emotion: An Essay in Aid of Moral Psychology*. New York: Cambridge University Press.
- Saint Augustine. c 400AD: 1953. *Confessions*. Translated by Vernon J. Bourke. Washington D. C.: The Catholic University of America Press.
- Scarantino, Andrea, and Ronald de Sousa. 2018. "Emotion." Edited by Edward N. Zalta. *Stanford Encyclopedia of Philosophy*. <<https://plato.stanford.edu/archives/win2018/entries/emotion/>>.
- Schacter, Stanley, and Jerome Singer. 1962. "Cognitive, Social and Physiological Determinants of Emotional States." *Psychological Review* (69): 379 - 399.
- Sellars, Wilfrid. 1991. *Science, Perception and Reality*. Atascadero, CA: Ridgeview Publishing Company.
- Singer, Peter. 2006. "Morality, Reason, and the Rights of Animals." In *Primates and Philosopher: How Morality Evolved*, by Frans de Waal, 140-158. Princeton: Princeton University Press.
- . 1993. *Practical Ethics*. New York: Cambridge University Press.
- Slingerland, Edward. 2003. *Effortless Action: Wu-wei as Conceptual Metaphor and Spiritual Ideal in Early China*. New York: Oxford University Press.
- Slote, Michael. 2015. "The Philosophical Reset Button: A Manifesto." *Dao* 14: 1-11.
- . 2010. *Moral Sentimentalism*. New York: Oxford University Press.
- Smith, Adam. 1790. "The Theory of Moral Sentiments." Library of Economics and Liberty. Accessed June 4, 2018. <http://www.econlib.org/library/Smith/smMS.html>.
- Smith, David Livingstone. 2011. *Less Than Human: Why We Demean, Enslave and Exterminate Other*. New York: St Martin's Press.
- Smith, Michael. 1998. "Emotivism." *Routledge Encyclopedia of Philosophy* (Taylor and Francis). doi:<https://www.rep.routledge.com/articles/thematic/emotivism/v-1>.

- Solomon, Robert C. 2004. "Emotions, Thoughts, and Feelings: Emotions as Engagements with the World." In *Thinking About Feeling: Contemporary Philosophers on Emotions*, edited by Robert C. Solomon, 134 - 155. New York: Oxford University Press.
- Stevenson, Charles Leslie. 1944. *Ethics and Language*. New York: Yale University Press.
- . 1937. "The Emotive Meaning of Ethical Terms." *Mind* 14-31.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109-166.
- Swinburne, Richard. 2001. *Epistemic Justification*. New York: Oxford University Press.
- . 1997. *Simplicity as Evidence for Truth*. Milwaukee: Marquette University Press.
- Tan, Monica. 2014. *Durian: love it or hate it, is this the world's most divisive fruit*. 1 October. Accessed August 27, 2019. <https://www.theguardian.com/lifeandstyle/australia-food-blog/2014/oct/01/durian-the-worlds-most-divisive-fruit>.
- Tappolet, Christine. 2018. "Précis of Emotions, Values, and Agency." *Philosophy and Phenomenological Research* 494-499.
- . 2016. *Emotions, Values, and Agency*. New York: Oxford University Press.
- . 2010. "Emotion, Motivation, and Action." In *The Oxford Handbook of Philosophy of Emotion*, edited by Peter Goldie, 325-345. New York: Oxford University Press.
- Taylor, Charles. 2003. "Gadamer on the Human Sciences." In *The Cambridge Companion to Gadamer*, 126-142. New York: Cambridge University Press.
- . 1985. *Human Agency and Language: Philosophical Papers*. Vol. 1. Cambridge: Cambridge University Press.
2006. *The Holy Bible*. Revised Standard Version Bible, Second Catholic Edition. Thomas Nelson Publishing.
- n.d. "The Sentences of Sextus." *The Gnostic Society Library*. Accessed May 3, 2020. <http://www.gnosis.org/naghamm/sent.html>.
- United States Holocaust Memorial Museum. n.d. *Anti-Jewish Legislation in Prewar Germany*. Accessed September 2019. <https://encyclopedia.ushmm.org/content/en/article/anti-jewish-legislation-in-prewar-germany>.
- University of Auckland. 2019. *Addressing Bullying, Harassment and Discrimination Policy and Procedures*. 18 February. Accessed August 29, 2019. https://www.auckland.ac.nz/en/about/the-university/how-university-works/policy-and-administration/human-resources1/employer-and-employee-responsibilities/prevention-of-bullying--harassment-and-discrimination-policy-and.html#par_pagedetails.
- van Roojen, Mark. 2018. "Moral Cognitivism vs. Non-Cognitivism." Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*. <<https://plato.stanford.edu/archives/fall2018/entries/moral-cognitivism/>>.
- Verbeek, Bruno. 2008. "Conventions and Moral Norms: The Legacy of Lewis." *Topoi* 27: 73-86.

- Virag, Curie. 2017. *The Emotions in Early Chinese Philosophy*. New York: Oxford University Press.
- . 2016. "The intelligence of emotions? Debates over the structure of moral life in early China." *Histoire intellectuelle des émotions, de l'Antiquité à nos jours* 16. <http://journals.openedition.org/acrh/6721>.
- Von Hildebrand, Dietrich. 1965: 2007. *The Heart: An Analysis of Human and Divine Affectivity*. South Bend: St Augustine's Press.
- . 1953. *Christian Ethics*. D. McKay Company.
- Vyasa, Krishna-Dwaifayana. c. 400 AD: 1890. *The Mahabharata*. Calcutta: Bhārata Press.
- West, E.W., trans. c. 250 AD: 1901. *Pahlavi Texts*. New York: Charles Scribner's Sons.
- Wheatley, Thalia, and Jonathan Haidt. 2005. "Hypnotic Disgust Makes Moral Judgments More Severe." *Psychological Science* 16: 780-784.
- Wong, David B. 2018. "Chinese Ethics." Edited by Edward N. Zalta. *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*.
- . 2017. "Moral Sentimentalism in Early Confucian Thought." In *Ethical Sentimentalism*, edited by Remy Debes and Karsten R. Stueber, 230-249. Cambridge: Cambridge University Press.
- . 2015. "Early Confucian Philosophy and the Development of Compassion." *Dao* 14: 157-194.
- . 2015. "Growing Virtue: The Theory and Science of Developing Compassion from a Mencian Perspective." In *The Philosophical Challenge from China*, edited by Brian Bruya, 28-58. MIT Press.
- . 2002. "Reasons and Analogical Reasoning in Mengzi." In *Essays on the Moral Philosophy of Mengzi*, edited by Xiusheng Liu and Philip J. Ivanhoe, 187-220. Indianapolis: Hackett Publishing Company.
- . 1991. "Is There a Distinction between Reason and Emotion in Mencius." *Philosophy East and West* 41 (1): 31-44.
- . 1991. "Response to Craig Ihara's Discussion." *Philosophy East and West* 41 (1): 55-58.
- Zagzebski, Linda. 2017. *Exemplarist Moral Theory*. New York: Oxford University Press.
- . 2010. "Exemplarist Virtue Theory." *Metaphilosophy* 41-57.
- . 2004. *Divine Motivation Theory*. New York: Cambridge University Press.
- . 2003. "Emotion and Moral Judgment." *Philosophy and Phenomenological Research* LXVI (1): 104-124.