DESIRING THE BAD UNDER THE GUISE OF THE GOOD

By Jennifer Hawkins

Desire is commonly spoken of as a state in which the desired object seems good, which apparently ascribes an evaluative element to desire. I offer a new defence of this old idea. As traditionally conceived, this view faces serious objections related to its way of characterizing desire's evaluative content. I develop an alternative conception of evaluative mental content which is plausible in its own right, allows the evaluative desire theorist to avoid the standard objections, and sheds interesting new light on the idea of evaluative experience.

I. INTRODUCTION

There is a familiar way of speaking about desire, as a state in which the desired object seems *good*, which ascribes an evaluative element to desire. I shall refer to it as *the evaluative conception of desire*. Many people find this view intuitive, or at least they find it a useful way of speaking. However, most current philosophy of mind rules out any such theory of desire. My aim here is to defend a version of the evaluative conception of desire, since I think the common way of speaking contains an important kernel of truth.[1] In addition, I aim to bring out something correct about *evaluative experience and evaluative thought* which is revealed by the evaluative conception of desire, properly characterized.

It will help if I distinguish my thesis at the outset from several others which I do not defend. If one assumes that all intentional action originates in desire, then one will naturally perceive a link between the evaluative conception of desire and what I shall call *the evaluative conception of intentional action*, the view that all intentional action is action taken *sub specie boni* or 'under the guise of the good'. Although I shall discuss both, I defend only the evaluative conception of desire. I remain agnostic about the evaluative conception of intentional action.

[1] My title invokes the titles of two previous papers: M. Stocker, 'Desiring the Bad: an Essay in Moral Psychology', *Journal of Philosophy*, 76 (1979), pp. 738–53; D. Velleman, 'The Guise of the Good', *Noûs*, 26 (1992), pp. 3–26.

This may seem inconsistent. But the appearance of inconsistency is only generated by an absurd, though extremely common, use of the term 'desire'. To maintain that *all* intentional action originates in desire requires a broad definition of 'desire', on which any 'pro-attitude' counts as a desire. For clarity, I shall refer to this broad conception as *desire$_W$*. The sense of 'desire' I adopt for this paper is not desire$_W$, but rather one much closer to the familiar pre-philosophical notion. As I understand it, desire only *sometimes* gives rise to action. In other cases, it is precisely desire that we resist when we act intentionally.

A second important disclaimer concerns terminology. While I have already admitted to finding certain uses of 'desire' counter-intuitive (e.g., desire$_W$), it is not my purpose to take a strong stand on the *meaning* of 'desire'. Rather, I shall indicate a set of familiar mental states that clearly fall within the ambit of the term 'desire' and argue that *all of these states* are evaluative. If anyone insists that other, non-evaluative, mental states have an equal claim to be included in the extension of 'desire', so be it. This insistence would undermine my preferred way of formulating my thesis, as a claim about *all desires*, but it would not undermine the thesis itself, which is about the psychological nature of a certain class of mental entities.

Finally, in what follows, I am concerned with the proper psychological characterization of evaluative thought, and not at all with the metaphysics of value itself. I assume that the account I offer is compatible with the full range of views on that issue.[2]

I begin by reviewing the traditional debate about the evaluative conception of desire and spelling out the problems it faces. I then develop my own alternative account of evaluative desire, which avoids the traditional problems, and I end by illustrating how, on my account, desire functions as one of the rudimentary elements of evaluative experience.

II. WHICH STATES ARE WE TALKING ABOUT, AND WHY THINK THEY ARE EVALUATIVE?

My thesis concerns the narrower pre-philosophical sense of 'desire'. But what exactly is this sense? A distinction originally introduced by Thomas Nagel may help here. Nagel famously divided the broad category of desires$_W$ into 'motivated desires' and 'unmotivated desires'.[3] The essence of the distinction has to do with reasons. Motivated desires are states or dispositions

---

[2] For a contrary suggestion, see M. Johnston, 'The Authority of Affect', *Philosophy and Phenomenological Research*, 63 (2001), pp. 181–214.

[3] T. Nagel, *The Possibility of Altruism* (Princeton UP, 1970), p. 29.

that we have *because* we recognize reasons for having them.[4] Unmotivated desires, by contrast, are states or dispositions which lack this basis in reasons. They are states with which we simply find ourselves.

Desires, in my preferred sense, fall in this second category. The concept of an unmotivated state or disposition seems essential to the ordinary pre-philosophical notion of desire. Not only is desire not something we reason our way *to*, but once desire exists it is generally not sensitive to reasons in the way other attitudes are. Desire none the less has the potential to motivate us, and sometimes we act simply because we desire something. Hence although not itself based on reasons, desire sometimes plays the psychological role of a reason for action. (Ordinary use of the phrase 'a reason for action' is ambiguous between (a) the notion of an apparent reason for action – a psychological condition in which people think they see a reason for acting, and (b) the non-psychological notion of that which actually supports or counts in favour of action. In this paper, I am exclusively interested in the first notion of *apparent* reasons for action.)

My claim is that all or most of these familiar unmotivated states, the kind that life just springs upon us, are evaluative. I say 'all or most', because some theorists might carve up the category of unmotivated desires even further, leaving room for some completely non-evaluative but still unmotivated states or dispositions. For example, some might claim that agents occasion-ally act on unmotivated *whims*, while maintaining that these are distinct from desires and are not in any sense evaluative. Or it might be claimed that unmotivated appetites such as thirst (though often accompanied by desire) are not, strictly speaking, the same as desire, and are not evaluative in them-selves. About these cases, I remain agnostic.

It is my agnosticism about such states that leads me to be agnostic about the evaluative conception of intentional action. I see no need to take a stand on whether all (as opposed to most) of the unmotivated motivators are evaluative. I also assume that the motivated motivators do involve evalua-tion, for these are precisely states that we come to have *for reasons*. In the account that follows, motivated desires turn out to be evaluative in a different sense from unmotivated desires. But if I am right, then many more intentional actions than current theory allows (even if not all) will turn out to be based in evaluative thought (of one sort or another).

What, then is the appeal of such a view? A full discussion of the merits of an evaluative conception of desire is beyond the scope of this paper. But while my primary intention is to describe a new and interesting way of

---

[4] Here I depart slightly from Nagel's own presentation. He describes motivated desires as those desires at which we arrive by 'decision and deliberation' (p. 29). See G.F. Schueler, *Desire: its Role in Practical Reason and the Explanation of Action* (MIT Press, 1995), p. 23.

developing the evaluative conception of desire that avoids certain traditional objections, a few remarks about the merits of the view are still in order.

The primary appeal of the evaluative conception of desire lies in the fact that desire often seems to play the psychological role of reason for action. However, if something is to play that role, it must have a certain kind of psychological structure or content. Moreover, this content must be capable of rationalizing action. It is not enough to view desire as a propositional attitude (plus the appropriate motivating tug), as most contemporary theories do. Unless its conceptual content is *evaluative*, it is unclear how desire can make sense of our actions in the way it seems to. Warren Quinn famously made a similar point by contrasting what it is like, from an agent's perspective, to act because one sees *something* good about the action chosen, as against what it is like simply to find oneself with an unintelligible urge.[5] Action of the second type feels mechanical and alien. One cannot even identify an apparent reason for why one acts as one does.

As we all know, action that flows from unmotivated desires is not like this. In virtue of the particular way a desired object enters awareness, action based on desire makes sense, in a way in which action based on the urges Quinn describes does not. This all too familiar feature of mental life needs to be accounted for. Describing desire as an evaluative state, one in which the desired object *appears good*, seems to offer the right kind of explanation. After all, pursuing something that strikes one as good has a kind of internal intelligibility.

III. PROBLEMS

The most familiar complaint against the evaluative conception of desire is that it collapses the distinction between desire and belief, turning desire into a species of evaluative belief. This is not simply a problem for Humeans. Even non-Humeans, who have no problem *per se* with the idea that beliefs can be motivating, continue to distinguish between desire and belief. A complete collapse of the distinction would have results that most theorists would find counter-intuitive.

The equation of desire with belief is counter-intuitive from both sides: it distorts our common notion of *evaluative belief*, as well as our common notion of *desire*. How it distorts *evaluative belief* is shown by a standard example of desiring the (believed) bad, originating with Gary Watson.[6] A mother is

---

[5] W. Quinn, 'Putting Rationality in its Place', in P. Foot (ed.), *Morality and Action* (Cambridge UP, 1993), pp. 228–55.

[6] G. Watson, 'Free Agency', *Journal of Philosophy*, 72 (1975), pp. 205–20, at p. 210.

overcome with exhaustion and irritation because her baby will not stop crying. As she bathes the child, she desires, in a moment of desperation, to drown the child in the bath water. If she were to do so, the explanation of her action would refer to her desire. But not all such cases are ones in which she *values* drowning the child. So we do not want to say that the mother has an evaluative belief that drowning the child would be *good*. The story is about being presented with a momentary temptation to do that which one *does not believe to be good*. It is precisely the fact that the mother does not value drowning the child which explains why it occurs to her only in a moment of desperation and exhaustion. It also explains why, even as she experiences the desire, she is simultaneously bothered by her own thoughts. But if we describe her desire as a state with evaluative content, and cannot find a way to differentiate that state from an evaluative belief, then we do seem to saddle her with an evaluative belief of the wrong sort. For ease of reference, I shall dub this woman *Frazzled*.

Several responses are available in defence of collapsing the distinction between desire and evaluative belief, but none is plausible. First, one might insist that Frazzled has both beliefs – the belief that drowning would be good (all things considered), and the belief that drowning is bad (all things considered). This is deeply unsatisfactory, since it turns a case of temptation into one of irrationality (holding contradictory beliefs).

A second, more promising, approach might emphasize that a thing can be recognized as good in some respects, but not others. Frazzled may thus believe that a particular *aspect* of drowning the baby is good, without believing either that all aspects of it are good, or that overall she ought to choose in accordance with her desire. (For example, her desire to drown the infant could be recast as the belief that the relief she would thus achieve would be good. She can accept this while still believing that the bad aspects of drowning far outweigh the good.) Yet again the cost of avoiding contradiction is phenomenological distortion. The whole notion that this is a case of *temptation* is lost on this way of construing it because the object of desire has been subtly shifted from drowning the baby to relief from the baby's crying. While this describes a possible case, it does not accurately describe the present case. Phenomenologically, I find it much more accurate to say that, *qua* desire, there is something *about drowning the baby* that seems good to Frazzled. Clearly, if we wish to leave room for cases in which the very aspect of the thing desired is the aspect that is believed bad, then we must either abandon the evaluative conception of desire or find a way for evaluative theorists to avoid collapsing the distinction between belief and desire.

How does collapsing the distinction between desire and evaluative belief distort our normal intuitions about *desire*? David Velleman, who emphasizes

this concern, worries ('The Guise of the Good', p. 3) that if we adopt the evaluative conception of desire, and with it the evaluative conception of intentional action, we lose an appreciation for the myriad, often quirky origins of intentional actions. His central concern is that if we describe desire as an evaluative state, all sorts of intentional actions will be mis-described as cases in which the agent *seeks out value*, when in fact nothing as rational or as purposive as that is really occurring.

Focusing on Velleman's worries about the evaluative conception of *desire* (rather than of intentional action), I grant him that there are cases where agents act on unmotivated desires, desires that just come to them and have no real tie to their evaluative beliefs, and that describing such cases as actions generated by evaluative belief seriously distorts them. This can be displayed by a case of silly, pointless activity (adopted from a children's book):

> Frances is a young badger in elementary school. She likes to follow a little ritual at lunch time (one she learnt from her friend Albert), called making her lunch 'come out even'. What does this mean? Imagine that her lunch box contains a sandwich, a pickle, a carton of milk, and a Chompo bar (favourite candy). Rather than eating each of these items all up, one after the other, Frances prefers to switch between them, eating a little of the first item, and then eating a little of the second, and so on. In addition, she tries to make it turn out that eating each item requires the same number of bites. Thus if it takes her twelve bites to eat her sandwich, she must adjust her bites of pickle so that it takes exactly twelve bites to eat the pickle (and similarly exactly twelve sips to drink the milk and twelve bites to eat the Chompo bar). Then the last bite of sandwich is followed by the last bite of pickle and so on, and her lunch has 'come out even'.[7]

Why does Frances do this? The best answer seems to be that the ritual just appeals to her. She enjoys it in her little way. It strikes her as a good thing to do at the time, even though she could offer no reasons for believing it has worth. To say that she *values* the ritual, or has some belief about its worth, as we seem compelled to do if desire is recast as evaluative belief, would seriously distort the case. Frances' ritual seems to be a good example of an action which Velleman would forbid treating as a case of earnest evaluative pursuit; to that extent, I agree. However, it also seems natural to say of Frances that in desiring to make her lunch come out even, she sees *something* good in her ritual. While we should not collapse desire into evaluative belief,

---

[7] Frances and her sister Gloria (introduced below) are characters from Lillian and Russell Hoban, *Bread and Jam for Frances* (New York: Harper Collins, 1964).

I think we should continue to look for a way to capture the kernel of truth in this natural description.

I have claimed that collapsing desire into evaluative belief distorts our ordinary sense of what evaluative beliefs are, and of what desires are. However, an analogy between perception and desire may appear to resolve these problems.[8] Perceptual states are mental states that represent the world as being thus and such. Perceptual states have content. But perceptual states differ importantly from belief states; for example, it is possible to perceive a stick as bent in water without believing that it is bent. Thus perception provides a model for mental states that have conceptual content without necessarily demanding the attitude of acceptance which partly defines belief.

This might seem to explain cases like that of Frazzled: we could say that drowning her baby presents itself to her as a good choice, but that she resists this presentation and *does not for a moment really believe the action is good*. However, promising as this approach may be, it faces an objection I have not yet considered. This third objection, also raised by Velleman (p. 7), depends on the nature of conceptual content. It seems that desires can be had by creatures, such as very small children, who do not have evaluative beliefs (or at any rate, evaluative beliefs of the right sort). To believe that something is good, one must possess the concept 'good'.

I assume that the general concept 'good' can be distinguished from the special concept of *moral goodness*, and that the former may be acquired earlier in life than the latter. Even so, Velleman's objection poses a problem. We commonly assume that very small children, including those too small to have even that basic all-purpose concept of goodness, desire things. As Velleman argues, it is presumably through training antecedently existing desires that we impart to children evaluative concepts like 'good'. Frances, who is in elementary school, clearly possesses the concept 'good'. But Frances has a younger sister Gloria, a toddler who can hardly speak yet. It is not so clear whether Gloria possesses this concept. But whatever we say about that, it is clear that she has *many* desires. Indeed, her myriad desires partly explain why she bothers Frances so much. So a defender of the evaluative conception of desire must not only find a way of distinguishing desire from belief, but must also find a way of distinguishing *the conceptual content of desire* from the *conceptual content of evaluative beliefs*.

Although the perceptual model alluded to above improves on the simple equation of desire with belief, it cannot solve this problem. Perceptual states are generally thought to have content in the same way as beliefs have. A

---

[8] A view like this is developed by D. Stampe, 'The Authority of Desire', *Philosophical Review*, 96 (1987), pp. 335–81. J.J. Thomson defends a similar view in *Goodness and Advice* (Princeton UP, 2001), pp. 26–40.

perception of 'the Chompo bar in the lunchbox' involves having the relevant concepts – the same concepts as the belief that the Chompo bar is in the lunchbox requires. If desire presents something *as good* in a manner analogous to perceptual presentations, it still follows that to desire something requires possessing the concept 'good'. It is not clear that everyone (e.g., Gloria) who desires things possesses this concept. Nor is it clear, even among those who do possess it, that the concept is involved when they are struck by unmotivated desires. Must Frazzled's desire to drown her baby somehow engage her concept of 'good'?

The conceptual content objection can easily be misunderstood. For example, if Gloria is so small as to lack the concept 'good', one might think her lack of other concepts will prove to be just as problematic. But this is unlikely. The kind of desire discussed in this paper requires, minimally, the ability to focus upon an object and have thoughts about it (or proto-thoughts, a notion introduced in §IV below).

So a newborn, for instance, whose sensory capacities do not yet even represent the world *as a world of objects*, cannot have desires in my sense. Yet this is over-determined. It is true not simply because newborns lack evaluative concepts, but because desires (in my sense) must have objects, and newborns lack the capacity to isolate and focus upon an object in the sense required.[9] (This is not to deny that newborns have drives and reflexes, or that we often loosely refer to these as desires. I only deny that there are any real psychological parallels.)

Unlike the newborn, however, it is natural to suppose that an older child like Gloria, whose world has acquired at least a simple conceptual form, is capable of desire in the familiar sense I have been discussing.[10] Admittedly, her *range* of desires may be narrower than the range of a mature thinker.

---

[9] Newborns can discriminate perceptual stimuli: see, e.g., A.J. DeCasper and W.P. Fifer, 'Of Human Bonding: Newborns Prefer Their Mother's Voice', *Science*, 208 (1980), pp. 1174–6. However, even the most permissive accounts assign concept-acquisition to no earlier than three months: see P.C. Quinn, 'Category Representation in Young Infants', *Current Directions in Psychological Science*, 11 (2000), pp. 66–70; G.L. Murphy, *The Big Book of Concepts* (MIT Press, 2002), ch. 9.

[10] I deliberately leave Gloria's age vague because theorists disagree about exactly when a child's world has assumed 'a simple conceptual form'. This debate is both empirical and theoretical, for it is a theoretical question what exactly we mean when we say that an individual possesses concepts, i.e., what capacities we associate with concept possession. Philosophers and psychologists typically differ on this point. In the next section, I introduce the notion of a proto-concept. This is a kind of representational capacity that is not as sophisticated as what philosophers often have in mind when they ascribe concept mastery. Developmental psychologists have none the less produced significant evidence for such capacities in pre-linguistic children (less than two years old) (see the references in footnotes 9 and 15). I take it that such capacities may be sufficient for the presence of desire. If this is correct, then a more precise statement would be that little Gloria's world has acquired at least a simple *proto-conceptual form*.

Lacking the concept 'detective', for example, she cannot form the desire to be the world's best girl-detective, as her sister Frances might. This is simple common sense: having no idea what detectives are, she obviously cannot form desires about them. But recognizing that her range of desire is restricted by her limited conceptual capacities is still compatible with accepting that she can form many other simpler desires. Hence the issue really does turn on a question about the possession of *evaluative* concepts.

At this point, there are several options open to defenders of the evaluative conception of desire. First, they might retreat to the claim that only mature desires are evaluative. This remains problematic, for it seems that desires in my sense are a distinctive kind of mental experience, a kind that is phenomenologically similar at various stages of our lives. Even though we become more sophisticated thinkers over time, and the objects of desire change, the experience of desire remains constant. Indeed, it is something about this experience of occurrent desire, an experience we apparently have in common with children, that initially leads one to describe desire as evaluative. Hence to give up saying that the desires of children are evaluative is to give up trying to explain the very feature that generated the account in the first place.

A second alternative is to try to establish that a child as young as Gloria already possesses the concept 'good', or at any rate that the concept is acquired at exactly the stage when desire becomes possible (whenever that is). This would require us to spell out the possession-conditions for the concept 'good'. Although I shall say more about these conditions later on, here I simply register my scepticism about this general strategy. Even if it is not large, I suspect there is a temporal or developmental gap between the point at which children become capable of desire and the point at which they have mastered the concept 'good'.[11] Moreover, lengthy reflection on the issues faced by the evaluative conception of desire has led me to conclude that the evaluative theorist should not hope to salvage the idea that desire requires evaluative concepts in the traditional sense. Rather, what is needed is an alternative way to describe the content of desire, one that allows us to bypass the issue of when exactly concept mastery occurs.

This is related to the intuition I hinted at in connection with Frazzled. Although the desires of small children like Gloria help to bring the problem into focus, it is not clear even in cases of adult desirers that desire is an

[11] The suspicion I record here is grounded in theoretical assumptions, not empirical ones. I suspect that the best account of concept possession, or the account most congruent with what philosophers (and hence the traditional objectors) mean, is going to require more capacities than a small child like Gloria possesses. Yet I also suspect that many philosophers, while denying that Gloria has the concept 'good', will none the less allow that she has desires.

experience that employs evaluative concepts, i.e., that desiring something involves classifying it in the very same way as one classifies things when one develops beliefs about their goodness. Desire appears to be a very different kind of mental phenomenon from belief. Hence if there is an alternative way to explain the intuition that desire is evaluative without ascribing mastery of the concept 'good', it seems to me that we ought to explore that path.

I suggest that there is more than one form of mental content, and that desire may have a particular type of evaluative content, even though it is not the same as the content of an evaluative belief. Moreover, this type of content may be developmentally prior to mature evaluative content. Hence it may be capable of explaining the desires we ascribe to young children.

IV. CONCEPTS, MATURE THOUGHT, AND SOMETHING THAT IS NOT QUITE EITHER

Having illustrated the problems that face evaluative accounts of desire, I shall now set desire temporarily aside. To set the stage for the account that follows, I shall briefly consider the philosophical notion of a concept. I shall then present an example of a mental capacity that while concept-like in certain respects, also falls short of all the marks of a mature concept. The example is drawn from the philosophy of perception. For this reason, it is limited in what it can show about desire and evaluative thought. Nevertheless, there are a few distinctive features of this example that can be generalized, and they will serve as the basis for the alternative account of evaluative content developed in the next section.

I begin, then, with concepts. For philosophers, concepts are first and foremost the building blocks of mental content.[12] They are delineated in terms of their distinctive contribution to the various attitudes in which they figure. The important feature of contentful mental attitudes is the way in which their particular content, in combination with other existing mental attitudes, sometimes *requires*, or in other cases simply *allows*, the development of new mental attitudes. I shall refer to these developments of new attitudes on the basis of old ones as *mental transitions*. Although mental life is characterized by a rich variety of mental transitions, philosophers have generally been most interested in those transitions that are *normatively re-quired*, i.e., inferences in the proper sense. Hence concepts have come to be

[12] To be precise, either concepts are the constituents of mental content, or at least a person to whom a certain mental content is ascribed must possess the concepts required to specify the content. This leaves room for a Russellian view on which the constituents of mental content are not concepts, but the individuals and properties thought about, but I ignore this for ease of exposition.

individuated in terms of their contribution to the normatively required inferences between mental states that include them. They are identified by their inferential role or by the particular inferences that are mandatory for all thinkers who possess the concept in question.

Although particular concepts are individuated by their inferential role, several other significant characteristics of mature conceptual thought are worth noticing for my purposes. First, mature conceptual capacities are associated with a capacity for *abstraction*. By this I mean both the simple capacity to think about objects in their absence (e.g., the familiar way in which the flow of occurrent thought is largely independent of the on-going flow of sensory experience) and the capacity to employ a variety of more abstract concepts (e.g., concepts that have little or no relationship to features of the world as perceived). Secondly, mature conceptual capacities are accompanied by a certain degree of self-awareness – in particular, aware-ness of the processes of thought itself. A mature thinker is a being capable of noticing the way in which new attitudes develop from old ones, and hence also capable of raising the question of whether a particular mental transition *makes sense*. In other words, the arrival of greater awareness comes with awareness of the normative constraints on thought, and with the capacity to stand back from one's mental attitudes and consider whether they are correct.

With this rough characterization of capacities for conceptual thought before us, I now consider an example of a kind of mental capacity that shares some aspects of conceptual thought, but lacks others. The example is drawn from the work of José Luis Bermúdez.[13]

Bermúdez is interested in what I shall call the *proto-conceptual* capacities of infants. In the example I borrow, he focuses on the process by which infants acquire the concept of a physical object. He begins by pointing to shifting views within empirical psychology about just how structured or unstructured the visual experience of small infants really is. Piaget and others of his generation thought that the concept 'physical object' was acquired at a relatively late point in infancy. Before that point, so they thought, the per-ceptual experience of very young infants is an 'undifferentiated chaos of noises, smells, patches of colour and flashes of light'.[14] However, more recent experiments by developmental psychologists suggest a different picture. The experimenters tested infants' awareness of physical objects by first habituating them to the presence of particular gerrymandered objects and then manipulating the visual appearance of these items, so as to make

[13] See J.L. Bermúdez, *The Paradox of Self-Consciousness* (MIT Press, 1998), chs 3–4.
[14] Bermúdez, p. 66, referring to J. Piaget, *The Construction of Reality in the Child* (New York: Basic Books, 1954).

the objects appear to behave in unusual, non-object-like ways (e.g., creating the appearance as of an apparently solid object passing through another one). The experimenters carefully recorded looks of surprise and shifts in the infants' attention patterns, taking these as evidence of the infants' sensitivity to objects *as such* and as evidence of the expectations that infants had of their sensory world. Indeed, the experiments seem to reveal a fairly sophisticated sensitivity to the presence and behaviour of physical objects.[15]

Bermúdez points out (p. 67) that *if* one accepts the experimental findings and the conclusions drawn about the perceptual world of these infants, then an important question arises about how to interpret these results. One possibility is to conclude that human beings acquire the concept 'physical object' much earlier than was previously thought. The other alternative is to acknowledge that while these infants have significant and previously unrecognized concept-*like* capacities, these capacities are not deserving of the label 'conceptual capacities'. Bermúdez defends the latter view on the ground that while the infants' capacities share certain features of mature thought, they just as importantly lack others.

The similarity lies in the fact that these infants appear already to have something that plays a limited inferential role in their mental lives. Through their reactions, they reveal a growing sensitivity to what Bermúdez (p. 72) calls 'object properties'. These are the properties that every mature thinker would recognize as being partly constitutive of the concept 'physical object', for example, the property of following a continuous trajectory through space-time, of having a determinate shape, of being impenetrable, etc. Sensitivity to these properties might seem sufficient for attributing the concept to these infants. However, Bermúdez (pp. 69–71) emphasizes that although the infants reveal a primitive sensitivity to the presence or absence of these properties, and even appear to engage in a limited number of inference-*like* mental transitions that respect these properties, they are still not capable of genuine inference. Genuine inference requires not only a disposition to make certain mental transitions, but also a rational sensitivity to the *correctness* of such transitions. This in turn seems to require the capacities for abstraction and self-awareness I mentioned before. Certainly, these infants lack these more sophisticated capacities.

For my purposes, Bermúdez's next suggestion is of particular significance. He goes on to suggest not only that perceptual states are structured or organized in proto-conceptual ways, but that these capacities may be a part

[15] See R. Baillargeon, 'Physical Reasoning in Infancy', in M.S. Gazzaniga (ed.), *The Cognitive Neurosciences* (MIT Press, 1995); E. Spelke and G.A. Van de Walle, 'Perceiving and Reasoning about Objects: Insights from Infants', in N. Eilan, R. McCarthy and M.W. Brewer (eds), *Spatial Representation: Problems in Philosophy and Psychology* (Oxford: Blackwell, 1993); E. Spelke, 'Principles of Object Perception', *Cognitive Science*, 14 (1990), pp. 29–56.

of a semi-autonomous mental system. On his view, proto-conceptual capacities are developmentally prior to conceptual capacities, and prior possession of such 'proto-capacities' may even be required in order to develop the mature capacities. But even though the mature conceptual capacities may develop from these simpler ones, he also suggests that the proto-capacities continue to function and play a significant role in mental life, even when the more mature conceptual capacities are in place. For example, while his account suggests that the concept 'physical object' emerges from the proto-conceptual object-sensitivity of perceptual experience, it also suggests that our visual experience *continues* to be structured in this proto-conceptual way, even once we become fully capable of thinking thoughts about objects. Thus, independently of our thoughts about objects, visual experience continues simply to present us with apparent objects.

As I emphasized earlier, this extended example from Bermúdez is only intended to serve as a jumping-off point for an account of a very different sort. Bermúdez defends a complex set of theses about perception, and it is not my intention to take a stand here on their plausibility. Whether or not his account holds true of perceptual experience, it offers a useful model for thinking about contentful mental states that may none the less differ significantly from the kinds of mature states upon which philosophers typically focus. Moreover, the idea that there may be an on-going relationship between more primitive ways of organizing our experience and our later, more refined ways of doing so is extremely suggestive for thinking about the relationship between affective states and mature conceptual thought.

V. EVALUATIVE MENTAL CONTENT WITHOUT
THE MATURE CONCEPT 'GOOD'

I shall now consider how the general idea of a much simpler set of mental capacities, embedded in a semi-autonomous mental system, may help to explain certain features of desire. I begin by reviewing what is commonly believed about *mature* evaluative concepts. According to one tradition, the all-purpose evaluative concept 'good' is an unanalysable term, a conceptual primitive. This means that while 'good' can be identified (like all concepts) by the unique inferential role it plays in mental life, it is not possible to explain this role in terms of the inferential properties of other simpler concepts from which it is composed. In essence, this is the view of 'good' which G.E. Moore made popular at the beginning of the twentieth century.[16] It is still widely influential.

[16] G.E. Moore, *Principia Ethica* (Cambridge UP, 1903).

There is an alternative tradition, however, according to which 'good' is not conceptually primitive. 'Good' can be defined in terms of the concept of *a reason*, or perhaps (with certain qualifications) in terms of the idea of *fittingness*.[17] On this view, to judge that an object is good is to judge that certain kinds of positive responses to it *make sense* (responses like *going for* it, *attending to* it, *preserving* it *from harm*, etc.). It is to judge that there is (at least a) reason for some such response. The notion of 'fittingness' provides an alternative way of stating the idea: something is good if certain positive responses to it are fitting or appropriate, i.e., if the responses 'fit' the object. However, the word 'fitting' has connotations which may not always apply. Often, when we say that a response is fitting, there is a suggestion that the response makes sense relative to a particular set of norms. But judgements about goodness are not always linked to a particular set of norms. So we may use the language of fittingness here only if we keep in mind that it merely indicates the presence of a reason.

If this analysis of 'good' is correct, then the notion of a reason turns out to be conceptually primitive. It is this analysis of the mature concept 'good' that I shall use in developing my alternative account of pre-linguistic evaluative content. I shall now consider the hypothesis that the mature concepts 'good' and 'reason for' *both* originate in a simpler type of *affective* experience, a type of experience humans become capable of well before they acquire the kinds of mature conceptual capacities moral philosophers typically discuss. I shall consider what reasons there are for accepting this kind of account only after I have finished spelling it out.

Importantly, this is a story not about perception, but about *affect*. Affect is not the same as emotion, even though affective capacities underlie our capacities for emotion. Our terms for emotions ('anger', 'joy') pick out particular eruptions or disturbances of affect, usually disturbances or eruptions that have already been conceptualized by the subject in particular ways. Hence, while our capacity for experiencing anger (say) is grounded in affect, anger itself is a state or disposition that emerges from the integration

---

[17] The analysis in the text is my own. While it can be seen as a natural outgrowth of a tradition of analysing 'good' in terms of other normative concepts, it is not itself one of the original analyses of that tradition. Examples of this tradition include H. Sidgwick, *The Methods of Ethics*, 7th edn (London: Macmillan, 1907), pp. 112–13; F. Brentano, *The Origin of Our Knowledge of Right and Wrong* (1889), tr. R.M. Chisholm and E. Schneewind (London: Routledge & Kegan Paul, 1969); C.D. Broad, *Five Types of Ethical Theory* (London: Routledge & Kegan Paul, 1930), pp. 277–8, 283; A.C. Ewing, 'A Suggested Non-Naturalistic Analysis of Good', *Mind*, 48 (1939), pp. 1–22, and *The Definition of Good* (New York: Macmillan, 1947), ch. 5. I am grateful to Tom Hurka for drawing my attention to Broad and to the relations between these writers. For the contrast between G.E. Moore and this tradition, see Hurka, 'Moore in the Middle', *Ethics*, 113 (2003), pp. 599–628. A recent addition to the tradition is T.M. Scanlon, *What We Owe to Each Other* (Harvard UP, 1998), pp. 97–8.

of affect and cognition. Anger acquires its particular subjective quality
because of affect, but anger cannot be explained in terms of affect alone.
Affect is a constant of conscious mental life in a way in which emotion is not.
Moreover, affective capacities are apparently functional well before mature
cognitive capacities. Although affect interacts with cognition once such
capacities are in place, much of affective life need not ever be given a
conceptual shape. Hence our capacity for affect is the capacity for a certain
type of continuous, subjectively distinctive, non-sensory responsiveness to
the world.

Philosophers have typically thought of affect as either a simple internal
push or pull, or perhaps as a kind of mental 'colouring' that, once concep-
tual states are in place, gives a particular subjective 'hue' to cognitively
structured mental states. What I suggest instead is that affect may itself be
primitively structured in ways that allow for affective presentations *as of* the
world being a certain way or having certain properties. These presentations
would not be conceptual, but at most 'proto-conceptual'. In short, what I
shall explore is the possibility that affect may present objects *as good*.

Returning to Bermúdez for a moment, he described a point in an infant's
life where it becomes sensitive to the presence of objects *as objects*. He then
suggested that from this point onwards, an infant's perceptual experience
may represent an object as having certain properties (e.g., the property of
being square) even though, as yet, the infant does not possess the concepts
of those properties (e.g., the concept 'squareness'). He calls the kinds of
mental attitudes generated by such experiences 'proto-beliefs' (pp. 96–100,
118). Significantly, proto-beliefs can influence the downstream shape of
mental life in ways reminiscent of mature beliefs. Belief typically gives rise to
inferences and new dispositions, and in conjunction with particular other
mental states or dispositions, it influences action. Similarly, in virtue of their
proto-predicative structure, proto-beliefs give rise to certain types of mental
transitions (which are inference-like), create new dispositions for reaction
(such as surprise, if the same object later appears not to be square), and in
conjunction with other mental states or dispositions, influence behaviour.

During this time, the infant is gradually moving towards concept mastery,
in this case, mastery of both 'object' and 'square'. But what is this process
like, and how are we to describe full concept mastery? Bermúdez (p. 75)
describes it as a gradual shift from 'implicit understanding' to a level of
'explicit understanding'. The infant moves from 'a cognitive level at which
understanding [e.g., that the object is square] is manifested in thwarted
expectations and reactive behaviour to a cognitive level at which the reasons
and grounds of those expectations and reactive behaviour are themselves
appreciated'.

Returning to affect, I now consider the possibility of a proto-predicative structure in affective experience. In the mental life of an infant or very young child, at a stage that occurs sometime *after* the perceptual world has resolved itself into a world of objects and their properties, I hypothesize the existence of a certain moment when an experienced sense of 'fit' emerges. As I imagine it, young subjects begin to experience certain of their own positive responses as fitting particular objects. It is important to keep in mind that, in this context, the word 'fit' is being used to gesture at a very simple kind of experience, a primitive feeling *as of* certain responses *making sense* or *feeling right*. Another way of trying to describe this experience is to say that certain responses have *subjective intelligibility* for the subject. The experience is proto-predicative because it is an emerging (and as yet very simple) awareness of, and feeling about, the relation between object and response. It seems plausible to imagine that some such experience is the most primitive experience *as of there being a reason* for responding to the world in a certain way. If this is right, then this offers an account of how a proto-concept of 'good' emerges from affective experience. I shall call these simple affective experiences of something as good 'evaluative impressions'.

A certain distinction from the philosophy of mind may help to clarify the nature of the kinds of occurrent mental states I am describing. In his book *On the Emotions*, Richard Wollheim distinguishes between the *intentionality* of an occurrent mental state (where he allows this term to take on a broader meaning than some philosophers admit) and the *subjectivity* of the state. On his view, intentionality is just a way of referring to the *content* of an occurrent mental state, that in virtue of which it secures its 'aboutness'. Subjectivity, on the other hand, refers to *what it is like* to experience a particular occurrent state. Wollheim insists that all (or almost all) occurrent mental states have both features. Of course, he acknowledges that some states owe their unique character more to one feature than to the other: what is distinctive about occurrent belief is its conceptual content and its resultant inferential consequences, whereas in the case of pains in particular parts of the body, the most distinctive aspect remains their subjective feel. Although they are frequently independent, Wollheim insists that there are often complex relations between subjectivity and intentionality. Indeed, there are even cases where the two 'fuse' so that it becomes impossible, at least at the level of experience, to separate them. An example from perception is Wittgenstein's famous duck–rabbit drawing: in this case it is virtually impossible to distinguish the shift in intentionality, the shift in what the drawing appears to be a drawing *of*, from the shift in subjectivity, the shift in how the drawing *looks*, or in *what it is like* to see it.[18]

[18] R. Wollheim, *On the Emotions* (Yale UP, 1999), pp. 6–8.

With this distinction in hand, I return to the states I call evaluative impressions. The *subjectivity* of such states is what it is like to experience one of them. One way to describe what I have called *subjective intelligibility* is to consider what the *absence* of such a feeling might be like. As discussed earlier, Quinn famously suggested an example in which a subject finds herself moved to grab an object, but in a way that, from the inside, feels mechanical and alien. This *quasi*-mechanical experience of agency, in which one recognizes that one is being internally pushed towards a type of action, even while failing to see *any* internal rationale for it, is to be contrasted with experiences in which a potential response (such as reaching for an object, or some other response) has a direct and obvious intelligibility. Certain responses just make sense, as it were, from the inside. Although Quinn was not discussing infant experience, I think he was describing a kind of affective experience we share with infants. The distinctive part of my claim is that in addition to their unique subjectivity, evaluative impressions have a primitive form of intentionality. Indeed, they secure this intentionality in part through their subjectivity. What emerges is a proto-thought *about* responses to objects, a thought which depends on the fact that certain responses to objects are subjectively experienced in a particular way as *making sense*. Moreover, like mature thoughts, these states have downstream effects on mental life.

Where in all of this does desire enter? As I see it, desire is one type of evaluative impression. It is only one type, since the word 'desire' is typically limited to impressions of objects which the subject does not yet have or actions the subject has not yet done. A person who does not have a kitten may desire to have one, or may desire to hold a kitten when not currently doing so. But evaluative impressions are not limited in this way. Evaluative impressions may extend to affective appreciation of what one has or of what one is doing, e.g., a sense that holding this kitten right now makes sense. Since desire is concerned with possibilities that have not yet been realized, it might be thought that this creates problems for the claim that young subjects have desires. However, while this does rule out infants who do not yet even perceive a world of objects with properties, it is unlikely to rule out young subjects capable of other types of proto-beliefs. From a very early age, infants perceive options for response (e.g., they recognize, in a primitive sense, that the object seen is one that could be grabbed, etc.). Once this simple capacity is in place, then we can say (borrowing a phrase from Mark Johnston, p. 186) that certain options just present themselves as 'default pathways'. The options are not presented as neutral; rather some potential responses appear (prospectively) to *make sense*.

This is the first part of my account, the part about evaluative impressions. But I am also interested in the idea, suggested by Bermúdez's account, that

certain proto-conceptual capacities are the springboard from which more sophisticated capacities develop. In the case of evaluative thought, I imagine that this next part of the story has at least two stages.

Bermúdez plausibly claimed that the move to concept mastery is largely the move from implicit to explicit understanding. Concept users understand why they classify certain things as instances of the property in question. What, then, might be the first stage in the acquisition of the concept 'good'? We can imagine a small child gradually gaining awareness of its own tendency to treat certain objects and responses as fitting one another. The child is then poised for the first stage of concept mastery. Suppose that, at first, those around the child largely restrict use of the word 'good' to those objects of which most of us have similar, positive affective experiences. For example, think of sweet things and imagine a parent saying, 'Now isn't that juice good?' (perhaps followed by the parent drinking some). Here, the parent is reinforcing an idea about the *relation* between the juice and drinking. There is a *fit* between certain properties of sweet fruit-juice and drinking the juice. The child can pick this up because the object labelled 'good' is already one it experiences as offering a reason. From the child's perspective, drinking juice already makes sense.

Stage two occurs when the child encounters others using the word to refer to objects that do *not* directly evoke the familiar responses. So now imagine a parent saying 'The broccoli is good' to a child who does not like the taste of broccoli. To the child's surprise, it discovers people apparently saying things about reasons where it experiences none of the familiar internal intelligibility of reasons. Still, having now grasped the connection with reasons, and the basic idea of a fit between object and response, the child is poised to grasp that the interlocutor is communicating something about fit and is recommending a certain response. Curiosity about what this fit could be will lead the child forwards into discussions of reasons. It is now on the cusp of being able to abstract the idea of a reason from the immediate experiences of default pathways, and becoming able to discuss with others what reasons there are. Over time and with experience, the child's appreciation of what reasons there are for what types of actions will expand significantly. Somewhere along this path, but only after becoming capable of thinking about reasons and goodness apart from its immediate evaluative impressions, the child will count as having mastered the concept 'good'.

This sheds light on Velleman's earlier claim that we teach children evaluative concepts by training their antecedent desires. On my view, there is a kernel of truth in this, in as much as children acquire the word 'good' because of the way we introduce it in relation to the things they desire. But whereas for Velleman this involves training utterly non-evaluative desires,

on my account such training is only possible because children already have experiences *as of* things being good. Hence a more accurate description of what occurs is that we expand children's range of evaluative appreciation, by building on the evaluative capacities they already possess.

On this account, mature evaluative content emerges from a simpler form of affective mental content. But both forms have a key structural similarity in virtue of which they both deserve the title 'evaluative content'. Both forms present subjects with a reason-giving fit between an object and a response, a fit that at one (pre-conceptual) level is *felt* as reason-giving and that at another (mature cognitive) level is *judged* to be reason-giving. But since, on my view, positing goodness just is positing the presence of such a fit, any level of mental content capable of presenting such a fit is naturally to be described as a content that presents something *as good*. If this is right, then we have a way of ascribing *evaluative* mental content to a small child like Gloria, a way that does not depend on whether she has yet mastered the concept 'good'. Whether she has mastered that concept or not, she is capable of experiencing things as good – of having evaluative impressions; and this is what we report when we report her desires.

There is one last but still significant part to this story. Just as Bermúdez suggested that our visual processing systems may continue to parse the visual array in terms of objects *independently* of the way we think about objects, so I suggest that affect may continue to present things to us *as good* long after mature evaluative concepts are in place. When I say that affect is a semi-autonomous system of mind, I aim to capture both the idea that affect interacts significantly with cognition (it is not utterly autonomous) and that it is none the less capable, at times, of conflicting with cognition.

The integration of affective evaluative impressions with ordinary thought is significant. In the ordinary course of life, we have a constant stream of evaluative impressions. Many of these we act on. Sometimes action is un-reflective. But in other cases, noticing how the object 'strikes' us, we pause to consider our own feelings and in so doing 'translate' the experience into conceptual thought. Imagine a woman working on a beautiful spring after-noon. Gazing out of the window at the newly green leaves, and sensing a particular fresh quality in the air that has been absent all winter, she suddenly desires to go for a walk. If she has no other obligations holding her back, she may simply walk out without further reflection. Imagine that someone questions her about her reason. At some point, the discussion is likely to bottom out in the claim from her that 'It just seemed like a good thing to do'. To explain her desire to someone else, she will have to give her experience a conceptual form, and one quite natural conceptualization of her earlier affective desire is *evaluative*. Such cases are important because

of what they reveal about the contributions of evaluative impressions to evaluative thought. For often such experiences will simply pass over into conceptual mental life without resistance or reflection. Hence this account offers a new way of making sense of the idea of evaluative experience, the idea that many (though not, of course, all) evaluative attitudes get started in something more basic.

However, the integrative process is not always so smooth, for conflicts can arise. Poor Frazzled has a desire for the (believed) bad. The thought of drowning her baby passes through her mind, and she is struck by a sense of the natural intelligibility of the move. Rather than passing through her mind fleetingly, the idea acquires impressive salience. It presents itself to her as the default pathway, the only thing she currently *feels* she has reason to do. Both the presence of this affective proto-thought and its power over her have a kind of *internal intelligibility* for Frazzled, even though another side of her is horrified by this fact. All of this is compatible with the claim that she does not believe, even for a moment, that drowning the baby is good, since she does not believe there is any *real reason* to drown it.

The power of this account is that it can explain Frazzled's desire without imputing to her an evaluative belief or even suggesting that her mature evaluative conceptual capacities have been engaged. Yet it reveals how, even so, one can continue to describe her state as evaluative. If (tragically) she were to act on her desire, we would be able to make sense of the idea that she acted on an apparent reason. On a lighter side, the account would allow us to explain what is evaluative about Frances' desire to engage in her lunch rituals (they strike her as primitively reason-giving), and it would allow us to ascribe desires to little Gloria.

If true, the account would vindicate the evaluative conception of desire. However, the remaining question is what reason there is for thinking the theory true. Part of the evidence has already been given, for it is phenomenological. We sometimes posit distinctions within mental life that, were they to exist, would explain significant *phenomenological* differences in aspects of mental life (as opposed to simply behavioural or functional differences). Of course, phenomenological evidence alone is not enough. Full acceptance of this theory would need much more empirical and conceptual work. It is important to ensure that the account meshes with empirical data, but also (since the empirical data will always underdetermine the account) that this theory really does serve to explicate widely held intuitions about evaluative experience and evaluative thought which cannot be captured in any simpler format.

In this paper, I hope to have done a fair bit of the work on the phenomenological side – at least enough to suggest that the theory is worth

exploring further. Moreover, I have demonstrated that the evaluative conception of desire is not incoherent, as some theorists have claimed. I have thus shown that the theory is a real possibility, and so worthy, given its intuitive status, of further investigation. No doubt further work will introduce refinements and alterations to the view as presented, but that is to be expected. It is offered in the spirit of illustrating how such an account might go.

VI. CONCLUDING REMARKS

I hope to have demonstrated that the evaluative conception of desire holds more promise than it is often credited with. I have offered an account which avoids collapsing desire into belief, as well as avoiding the view that only creatures with sophisticated evaluative concepts could have desires. Yet the account manages to retain the idea that there is something about the experience of desire that deserves the label *evaluative*. It thus succeeds in explaining how unmotivated desires could play the psychological role of reasons for action.

   In the process of defending this conception, I have introduced a new (sketch of a) theory of *evaluative* mental content. This theory clearly needs more work, but it ought to receive that work because it holds out interesting possibilities that extend beyond our interest in desire. This brings me back to my opening claim that, properly characterized, the evaluative conception of desire reveals something important about evaluative experience and evaluative thought. For my characterization of evaluative desire gives new meaning to the idea of *evaluative experience*, making such experience possible in a new and original sense. Moreover, it suggests that evaluative experience may be an important source of evaluative belief in a way not often allowed for in moral psychology.[19]

University of Toronto