Reply to Critics

TIMOTHY SCHROEDER *University of Manitoba*

Three Faces of Desire takes up the question of desire in mid-sentence, as it were. It does not begin with foundational issues, it does not defend first principles, it does not clarify the terms of the debate. Instead, *Three Faces* jumps straight into the question of what it is to want something, bringing to bear common sense, philosophy, and neuroscience in equal measure. But, as all my critics have reminded me in one way or another, questions about foundations and first principles are still central to philosophy and require as much care as the questions that *Three Faces* puts at the centre of its attention. In this reply, I will start with foundational issues, and then say a little more about each of the three faces of desire (reward, pleasure, and motivation) in turn.

1. Are Desires Propositional Attitudes?

To begin at the very beginning: Paul Thagard holds that desires are not all propositional attitudes; some are, but many are too simple for that. Ronnie de Sousa, meanwhile, holds that desiring that P is underspecified, since one may desire that P with different aspects, i.e., as a goal-oriented process or as an ongoing activity. Both thus worry at my most basic assumption about desiring, that desiring is a matter of desiring that P, where "P" stands for a propositional content. However, I think that a sensibly ecumenical understanding of propositional attitudes makes room for the phenomena mentioned.

Thagard's challenge is the more pointed. He reminds us of various phenomena that suggest—some rather strongly—that desire cannot be a propositional attitude. A wide range of non-human animals have desires (on my view, at least) while only a tiny range of non-human animals have

Dialogue **XLV (2006), 165-74**
© 2006 Canadian Philosophical Association /Association canadienne de philosophie

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

significant conceptual sophistication. I can desire that I have an apple in my possession, but how can an animal have that same desire without concepts such as SELF and POSSESSION? There are narrow ways of thinking about propositional contents on which this question would be very difficult to answer. My own way of thinking about propositional contents is broader, however. I think of propositional contents as individuated not by the concepts their entertainers deploy, but by how things must be in the world for them to be satisfied. When a lizard desires a fly, the way things must be in the world for the lizard's desire to be fulfilled is that the lizard catch and consume the fly. At the very least, this is the condition that would normally contribute to a reward signal in a fly-eating lizard (I imagine). The lizard missing the fly, or merely being landed on by the fly, or having the fly tied to its neck by a slender string, would certainly *not* trigger a reward signal. And, one way or another, the lizard is able to detect the fly-catching-and-consuming state of affairs and distinguish it from the others, and this is why the former causes a reward signal and not the latter. Hence, one correct way to characterize the lizard is by saying that it desires that it catch and consume the fly. Its ability to entertain such an attitude while lacking concepts such as SELF and CONSUME is a puzzle, but there are theories—those of Fred Dretske and Ruth Millikan come to mind especially—on which such contents can be entertained even by creatures that, intuitively speaking, lack the concepts that we would use to form thoughts with the same content. So I leave the details to such philosophers.

De Sousa's challenge can also, I think, be accommodated. It might be the case that for one person, representations of having sex are connected in a positive fashion to the brain's reward system (so that sex itself is motivating and pleasing), while in another person only representations of orgasm are so connected (and the enjoyable arousal of sex is a result entirely of representations of the approach of orgasm). In the two cases, it seems as though the contents of the desires are different, and differ exactly as process differs from activity. What my characterization of the contents of the desires has done is to move the linguistic expression of that difference from grammatical aspect alone to grammatical aspect plus noun choice ("having sex" versus "orgasm"). What the grammar of the brain is like is something about which I have no opinion, but since there are multiple ways of expressing exactly the difference de Sousa has in mind, and since our minds can clearly respond to these differences somehow or other, I have no particular worries that mental representations include aspect in one fashion or another, and so that desires do too.

What I do have a concern about is de Sousa's proposed principle, that "an object of desire should be desired in the temporal aspect most suited to its nature as a source of pleasure or satisfaction." Take a desire to care for a dying loved one. If this is thought of as a finite process, a readily completed task, that might take a significant emotional burden from one's

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

shoulders. Actually giving care is something one can do—show up at the hospital, say, fluff the pillows, hold a hand, and so on—and then one can leave in the knowledge of a job well done. If it is thought of as an ongoing activity, by contrast, then every moment at which the dying loved one suffers is a moment at which one's desire is frustrated. The emotional burden undertaken in having the second sort of desire is much greater. But those facts alone do not convince me that it is more rational to desire to care for a loved one as a process than as an activity.

2. Are Desires Natural Kinds?

My assumption that desires are propositional attitudes is only barely defended in *Three Faces*, but my initial assumption that desires are natural kinds eventually gets a whole chapter to itself (chap. 6). Even so, I have clearly not answered the critics, for Latham and Brook have points to press from one direction, and de Sousa—ever the joker in the pack—has points to press from another as well.

To be clear at the outset, RTD—the reward theory of desire—is something I hold to be a necessary truth that is known only *a posteriori*. But Noa Latham and Andy Brook think that we know enough *a priori* about desire to rule RTD out of contention. Latham's view is that there are behavioural dispositions sufficient to guarantee the presence of a desire. Brook's view is that the feeling of desiring—of longing, of having urges, and so on—is sufficiently important that no one can desire that P without having some tendency toward these feelings.

My general strategy in *Three Faces* is to emphasize how little can be explained by a theory of mind that builds something into the nature of desire, rather than holding it to be a causal effect. The fact that water is H_2O allows its transparency to be explained by its nature, allows its liquidity at room temperature to be explained by its nature, allows its tastelessness to be explained by its nature, and so on. But if water were nothing other than a transparent, liquid at room temperature, tasteless, etc., substance, then none of these explanations would be anything but trivial. Similarly for desires.

Latham takes up the challenge by pointing out that desires can be invoked in a number of non-trivial explanations of motivation, even if one holds there to be a metaphysical connection between desire and motivation. So, if I am motivated to buy a new car, it is non-trivial even on an interpretative account of the mind to say that I was so motivated because I desired a boost to my ego, which I thought the new car would provide. After all, it might rather have been a desire to have safe transportation, and the thought that my old car had become unsafe, that led to the purchase of the new car. This is a reasonable point. But what cannot be said without triviality is that I display the whole syndrome of desiring a new car because I desire a new car (plus the relevant facts about belief, habit,

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

and so on). To display the whole syndrome just is to have the desire, and so cannot be explained by it. To my mind, this is just as unsatisfactory as being unable to explain why water displays the syndrome of being watery by appeal to the fact that it is water (plus the relevant facts about light, liquids, and so on).

Likewise, Brook's view that a necessary part of desire is the feeling of wanting makes such feelings inexplicable in terms of desire. There can be informative explanations of why I have a feeling of craving in terms of my wanting this rather than that, but to say that since I want something very much I am subject to feelings of craving and longing is to return to the domain of the trivial. Absent such feelings (systematically, at least), I would not want the thing in question, on Brook's view. Once again, I am simply not subject to the same intuitions.

De Sousa's suggestion, that no one face of desire is necessary, but that some complex of faces might be sufficient, is—I think—the same as what I call the "mix-and-match theory of desire" in *Three Faces*. But, if the mix-and-match theory is correct, then we have an explanatory disaster: nothing a desire does can be explained in terms of desire because having a desire is having anything that does anything rather desire-like. The instinct to call something "a desire" whenever it much looks like a desire is saved, but at the cost of losing explanatory force.

Moving from general points about explanation to some of the nitty-gritty: Brook asks "Could a desire currently active in shaping my intention-forming and decision-making not be like anything to have?" His answer is that perhaps it could, "in a few oddball cases, but not in general." But it seems to me that most of our deliberation and conversation is shaped by what Hume called the "calm passions," by desires that are efficacious without there being anything it feels like to have them at work. As I work on this response, for instance, I am constantly deliberating on what to say. My desires to learn what is true, to be polite, to be thought well of, to express myself clearly, and so on are all clearly enough at work, given how my thoughts veer now one way, now another. Yet I do not feel anything like a craving to know the truth, or an urge to sound respectful. Similarly, in a conversation about philosophy very similar desires are in play, but the phenomenology is equally bland. Occasionally a situation is so significant that my phenomenology will include a feeling that I crave attention, or that I have an urgent need to be polite now, or something of the sort. But this is the exception. Consider also the desires that are at work regulating the driving of a car, and how typically only the threat of being late or in an accident brings any of them to the phenomenological realm. These sorts of considerations make me think that the phenomenology of craving, of urges, and the like cannot be typical of even occurrently influential desires.

In parallel, Latham wonders if anything could exhibit all the behavioural signs of desire without desiring. Here it is harder to come up with

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

an intuitively plausible counterexample, but consider a HLUT machine, an apparently artificially intelligent android that is simply driven by a Humongous Look-Up Table. (The idea is not original to me, but as I learned the acronym from its currency on the Internet circa 1994, I am afraid I do not know whom to credit). A HLUT machine has all the behavioural dispositions of an intelligent person, but this is simply because it contains an unspeakably huge look-up table that tells the machine, for every series of contingencies, what to do. The machine is, in effect, a giant list of conditionals, each of the form "if the world so far has been X, then do Y," with one conditional for every possible contingency. There is no way to build a HLUT machine smaller than the entire universe, I imagine, but that is no barrier to its conceptual possibility, and so to investigating the *a priori* knowable nature of desire. To most people, the obvious position to hold is that a HLUT machine does not have mental states. I side with the popular opinion. I am quite ready to agree that, *a priori*, there is little to say about what the origin of our apparently desire-driven behaviour must be like in order for it to really come from desires, but one thing at least can be said: it must not look like a look-up table! So I doubt there are sufficient behavioural conditions for the possession of desires.

Stepping back again, let me say something not appearing in *Three Faces* but which I think is true. Most of our kind terms are, by default, would-be natural-kind terms until we learn better. Consider the terms for the elements: "earth," "air," "fire," and "water." These were, I think, default natural-kind terms. If there had been natural kinds corresponding to all four, then all four terms would have been natural-kind terms. As it happened, only one was a term for a natural kind, so only one term is in fact a natural-kind term ("water"). I think of our terms for mental states as similar. Not every mental term names an actual natural kind, for there are no natural kinds available to be the referents for some terms. But when there is a natural kind, then the mental term names it by default, for by default when we name things we take ourselves to be carving nature at the joints.

This brings me back to de Sousa, who reminds me of how poorly species resemble natural kinds, and asks why I think desires are any different. Desires are different, I think, exactly because desires are not like species. Species are the sorts of things that are constantly turning into other species, continuously, without sharp branch points. Desires, on the other hand, just happen to be things that do not, in fact, have a plethora of intermediates between them and other mental states. Nature has, it seems, stuck to a particular pattern and conserved it, defending it against gross modification. This is the sort of natural kind that nature generally provides, above the level of subatomic physics: a kind that could in principle, but does not in fact, present a rich penumbra of borderline cases that blur into other, distinct kinds. Our cognitive systems do their best to lock onto these kinds, and, when they succeed, our concepts come to be of them.

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

So much for foundational issues. What about the details of the theory? I will begin with reward.

3. More About Reward

Inevitably, *Three Faces* engenders a certain amount of confusion and a certain amount of disbelief around its use of the term "reward." So what do I really mean by the term? And how, exactly, are desires related to rewards?

I mean by "reward" exactly what a teacher means when announcing that there will be a reward for the student who wins the spelling bee, and I mean by "punishment" exactly what an editorialist means when denouncing the dearth of serious punishment for white-collar crime. A reward is a wanted state, event, or the like (the ontology is not important) one gets as a result of doing something; a punishment is a correspondingly unwanted state or event. So rewards and punishments are, generally, things outside the head. But what makes them rewards and punishments are facts about desires—things inside the head. To learn about desires, then, one need only learn about the effects of rewards, for the psychological effects of getting a reward then turn out to be the psychological effects of having a desire satisfied.

Latham is sceptical that I really mean "reward" as a mental term. In a sense he is right, since rewards such as cookies are not mental at all. But he also thinks that therefore my talk about the reward system is, at most, talk about implementation. This, I think, is mistaken. What I call "the brain's reward system" is the system in the brain that makes it true that something counts as a reward for one. That is, it makes it true that one responds to a reward in the way that is characteristic of psychological responses to rewards (including overt behaviour, pleasure, changes in behavioural dispositions, changes in psychological dispositions, acquisition of new intrinsic desires, and changes in sensory sensitivities). I could have called this the R system. Had I done so, I could then have discussed the brain's implementation of its R system *via* dopamine, neural signals, and so on—perhaps this would have been clearer as to what is implementation and what is being implemented. The R system is itself psychological, insofar as we think that taking the right attitude to something constitutes it as a reward. The only question is what psychological state the R system is. The upshot of Chapter 2 of *Three Faces* is that the R system is desire. So we can think of desiring that P using two phrases, "desiring that P" and "having a representation that P set up to contribute to a triggering of a reward signal (i.e., a certain kind of learning)," and the point of *Three Faces* is to make the case that these two phrases are phrases for the same thing.

Brook is sceptical that rewards can be distinguished sharply from pleasure. But here the details seem very important: getting a reward does so much, normally. It guides behaviour, it makes one smile, it causes unconscious learning of all sorts. Pleasure, so far as empirical science has determined, makes little or no contribution to each of these things in many

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

cases, and the main contribution is always made by the reward system. Pleasure is causally distinct from the reward system, downstream of it. Why, then, hold that its involvement is partly constitutive of the fact that something is a reward?

What, now, is the relation of reward to desire? Brook wonders if it might be that by desiring we learn what is rewarding. Latham wonders if non-mental creatures such as plants might instantiate reward systems. These questions make it clear that I could be still more careful in saying how I link reward and desire, so here is the attempt: to desire that P is to be such that, when one represents that P, one tends to respond in the way constitutive of being rewarded, i.e., with a certain sort of learning ("contingency-based," or "reward" learning, as it is known). Because a reward system includes, as a part, representational capacities, a non-mental creature cannot be rewarded. A creature can only be rewarded when it can perceive or conceive of it being the case that P, for this initial perception or conception is required as part of the reward system. (This, I hope, answers Latham.) Because a reward system essentially works to drive learning, it is indirectly a source of learning about what is rewarding, but this is a very indirect sense. What we mostly learn from the activity of a reward system is not facts but dispositions: to act in certain ways, think in certain ways, associate certain ideas, treat states of affairs as inherently motivating, perceptually discriminate in certain ways, and so on. Any learning of facts about, for instance, what is rewarding is strictly secondary. (This, I hope, answers Brook.)

4. More About Pleasure

I count myself fortunate that none of my critics has taken issue with the phenomenological claims I make about pleasure, not least because I do not know of much other than an incredulous stare to use when someone tells me that pleasure is not a distinct feeling. But it hardly follows that no questions are raised. Three stand out here.

De Sousa has a pair of questions to address. One is a friendly question: what do I think about the idea that the pleasure of anticipation should be proportionate to the anticipated pleasure? The other is more tinged with scepticism: on my view, pleasure does so little that it begins to seem epiphenomenal. Can that really be right?

In response to the first: I am not convinced that de Sousa's principle is correct. I hold a slightly different (I think) principle: that in an epistemically fortunate animal (one not subject to illusions or hallucinations), the pleasure of anticipation will generally be proportionate to the pleasure of consumption. So, the pleasure I get from looking forward to eating a Montreal sesame bagel stands in a similar proportion to the pleasure I get from eating the bagel as the pleasure I get from looking forward to finishing a big pile of grading stands to the pleasure I get from actually finishing the pile, and so on. But de Sousa's principle, if I understand it correctly,

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

is that the pleasure I get from looking forward to the bagel or the completion of grading should be proportionate to the pleasure I believe I will get from the bagel or from finishing grading. And this I do not accept, because I argue that the capacity for pleasure is a kind of perceptual faculty, and because one does not demand that one's perceptual faculties and one's beliefs agree in general. After all, one's eyes may say there is nothing in the jar, while one's belief is that air remains in the jar, and there is nothing to fault in either the perceptual or the cognitive system. Likewise, I might have significant anticipatory pleasure from the prospect of meeting someone for a first date, while intellectually knowing that it is much more likely, given the evidence, that the date will be dreary. There need be nothing to fault in the mismatch between the more perceptual system—the felt pleasure of anticipation—and the more intellectual system—the belief about what pleasure lies to come. But I would certainly agree to the weaker claims that there are occasions on which one's beliefs are right about future pleasure while one's hedonic system is wrong (and vice versa), and that a reflective person who notes a mismatch ought to be careful about drawing conclusions.

In response to the second: I confess that I was surprised when I learned how unimportant a role pleasure plays in guiding behaviour. But eventually I lost that sense of surprise. Pleasure, it turns out, is in many ways like other states of consciousness such as the state of smelling cedars or the state of tasting fine chocolate. It is something we generally want, and which contributes to learning and motivation just as the smell of cedars or the taste of fine chocolate might. Its distinctiveness is in the fact that (almost?) everyone desires pleasure. But then, on my view, pleasure represents a net positive change in desire satisfaction, and such change brings with it a signal in the form of dopamine that tends to inculcate the formation of new intrinsic desires. So, pleasure is generally co-instantiated with a signal that tends to inculcate desire. So, pleasure is the state of consciousness that, above all other sensory states, tends to be desired, and desired strongly. Because pleasure is desired, and desired strongly, it is far from epiphenomenal, but because it is only one thing desired out of many, it is not nearly as powerful in our mental lives as it has been thought to be.

This talk about pleasure as representing change in desire satisfaction draws Latham's attention. He wonders if I have not taken on the hard problem of consciousness by giving an account of what it is to be pleased in terms of having a given representational content. In all modesty, I have to admit that this is already the theory of consciousness presented by Fred Dretske and Michael Tye, amongst others. But I do not think I require that particular approach ("representationalism" or "intentionalism") to be correct. All I require is that, necessarily, pleasure have the content I attribute to it. In the same way, someone such as Ned Block can hold that nothing could be qualitatively identical to my current state of visual con-

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

sciousness without it having the content that there is a flat, silvery surface about forty centimetres in front of me, while holding that there is something beyond that content that gives my consciousness its phenomenal quality. I make little of the possibility of a pleasure *quale* the feel of which is not identical to any representational content simply because positing one makes no difference to me either way.

5. More About Motivation

Finally, a few more words about desire and motivation. Much has already been said about my attempt to sever desire from motivation in ontology, and to reattach them only with causal ties. But there remain two other points I would like to take up.

De Sousa proposes to think of vice as being unable to stop choosing what no longer brings pleasure. One of the things I like about the view of desire in *Three Faces*—and one of the things lost if one ties desire too closely to motivation—is that it can explain de Sousian vice in plausible terms. Take smoking as an example. As I understand it, it is common for smokers to enjoy the first n cigarettes of the day, but as nicotine builds they find smoking less pleasant, and at a certain point enough nicotine has built up in the body that smoking is actually unpleasant. Nonetheless, many smokers find themselves smoking the $n+m^{\text{th}}$ cigarette while experiencing displeasure. Such choices are made in the absence of expected overall pleasure, and the absence of actual overall pleasure. (Related things can be said about some illegal addictive drugs as well.) Why? Investigation of the brain shows that nicotine contributes to the production of a reward signal quite independently of how one represents the world. That is, nicotine need not be desired in order to trigger a reward signal, because it hijacks the brain's reward system at its midpoint. One obvious result of such triggering of the reward signal is the inculcation of strong behavioural tendencies, tendencies that exist independent of desires. The result: choosing against both pleasure and desire, because of past behavioural learning induced independently of both pleasure and desire. The learning is independent of pleasure because pleasure is not required to trigger the reward signal; rather, it is a typical causal consequence of the signal. The learning is independent of desire because no representation gets activated to cause the learning. And because desire and pleasure are both causally linked to, but independent of, motivation, choice, and action, it becomes possible to choose that which is not wanted and does not please. De Sousian vice is thus made possible.

Thagard draws out the connections and disconnections between *Three Faces* on motivation and his and Wagar's GAGE theory. A few critical remarks will have to come before the praise, but I do find this convergence of views heartening. Wagar and Thagard put more emphasis on the amygdala and the hippocampus than I think is entirely warranted. As for the amygdala, massive damage to this structure (known to be crucial for

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press

the classical conditioning of fear responses, at least, and more besides) does less to impair motivation in general than one would expect on their model. As for the hippocampus, massive injury to the hippocampus generates retrograde amnesia (the familiar sort of amnesia), but amnesia is no barrier in principle to behaviour. The GAGE theory needs broader connections to the brain's conative and cognitive systems in order to be fully plausible. Yet (and here I can only suggest that the reader confirm this by direct contact with that article) there is also some wonderful depth of detail in parts of the GAGE theory that *Three Faces* can only envy. The computational modelling that backs GAGE is the sort of thing that will one day reveal just how desires lead to learning, to choice, and to pleasure in all the intricate ways that they do.

6. Summary

By Latham's reckoning, *Three Faces* awards its theory of desire a decisive pair of victories over competing theories (by 8-3 and 7-2, apparently), but by Brook's reckoning *Three Faces* suffers a defeat at the hands of common sense (3-1 against). One cannot help but feel that when the author awards the victories and the critic awards the defeat, this is in some way telling! But I am delighted to have lost, if that is really how things come out, to such a wonderful group of critics.

The central claim of *Three Faces* that remains most controversial at this point, even in reasonable company, is that desires should be thought of as ontologically independent causes of their most familiar associated phenomena. *Three Faces* takes phenomena such as motivation, behaviour, pleasure, and felt urges to act to be the causal effects of having desires, while Brook, de Sousa, and Latham all expressed the thought that this went too far. I have done something here to motivate the idea that desires are known by their effects, not by their natures, but in the end this idea's justification depends on other issues—considered in theories of explanation, semantics, mental causation, and so on—that will not be quickly resolved.

The central claim of *Three Faces* that remains least controversial at this point is the three-way causal interaction between, and ontological independence of, reward learning, motivation, and pleasure. De Sousa and Thagard have especially remarked on this as an interesting feature of *Three Faces*, and I am delighted. This three-way independence, and the patterns of causal interaction described in *Three Faces*, are data that will not go away, and will need interpretation of one sort or another. If the interpretation given in *Three Faces* is not satisfactory, then some other interpretation is needed! I look forward to seeing what it might look like, but I have a sneaking suspicion that it will look a lot more like the reward theory of desire presented in *Three Faces* than it will look like anything else philosophy has said about desire up until this point.

https://doi.org/10.1017/S0012217300000408 Published online by Cambridge University Press