Australasian Journal of Philosophy

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rajp20

Sympathy and subjectivity

Peter Carruthers [a]

[a] University of Sheffield

Version of record first published: 02 Jun 2006.

To cite this article: Peter Carruthers (1999): Sympathy and subjectivity, Australasian Journal of Philosophy, 77:4, 465-482

To link to this article: http://dx.doi.org/10.1080/00048409912349231

Australasian Journal of Philosophy
Vol. 77, No. 4, pp. 465–482; December 1999

SYMPATHY AND SUBJECTIVITY

Peter Carruthers

I. Introduction and Background Assumptions

This paper makes four main assumptions—two about consciousness and two about value—which together raise an important question concerning the possible scope of morality and moral concern. I shall make little attempt to defend these assumptions, beyond saying just enough to explain and motivate them. Each of them can be made to seem plausible, at least; and three have been defended at length elsewhere.[1] If these premises are granted, then fundamental questions are raised about the moral status of non-human animals. But even those who are not prepared to grant one or more of my assumptions should take an interest in the questions they generate, and in my proposed answers. For the discussion will lead us to look quite closely at the nature of psychological (as opposed to biological) *harm*,[2] and at the proper objects of sympathy; raising issues which have not, I think, been discussed before. Even if one or more of my assumptions are rejected, there will remain the question whether, and in what sense, sympathy is necessarily tied to subjectivity—which is not a question with quite the same practical import as the one I address, admittedly; but an interesting one nonetheless.

I.1. Assumptions about consciousness

Here are the two main assumptions about consciousness laid out in summary form, together with the initial conclusion which they entail:

Assumption 1: All of the mental states of non-human animals (with the possible exception of the great apes) are non-conscious ones.

Assumption 2: Non-conscious mental states lack phenomenology, or subjective 'feel'.

Conclusion 1: The mental states of non-human animals lack phenomenal feels (unless those animals are apes, perhaps—I drop this qualifier in what follows, for the sake of simplicity).

[1] For detailed explanation and defence of Assumptions 1 and 2, see my *The animals issue: moral theory in practice* (Cambridge: Cambridge University Press, 1992), chs. 6 and 8; my *Language, thought and consciousness: an essay in philosophical psychology* (Cambridge: Cambridge University Press 1996), chs. 5–7; and my 'Natural theories of consciousness', *European Journal of Philosophy* 6 (1998), pp. 203–222. For a classic defence of Assumption 3, see J. Mackie, *Ethics: inventing right and wrong* (London: Penguin Press, 1977), Part One.

[2] Any living thing which can be *damaged* (including plants and micro-organisms) can be harmed in the biological sense; but it is doubtful whether this notion is of any direct moral significance. I assume in this paper that the proper object of sympathy is *psychological* harm, involving either pain or the frustration of desire.

The conclusion C1 then generates, quite naturally, a further question. This is the main question to be answered in this paper, and is as follows:

> *Question 1:* Given C1, ought we to conclude that sympathy (and other moral attitudes) towards the sufferings and disappointments of non-human animals is inappropriate?

In my *Animals Issue*, chapter 8, I argued tentatively for a positive answer to this question. But I am now not so sure. Indeed, the main burden of this paper is to demonstrate that there is a powerful case for answering Q1 in the negative (and one which does not need to rely upon an objectionable realism about value—see A3 below).

Why should we think that animals don't have conscious mental states? The short answer is that some or other form of Higher-Order Thought (HOT) theory gives the best account of mental-state consciousness.[3] And although there is vigorous debate about whether chimps and other apes are capable of HOTs,[4] it is generally agreed that cats, dogs, sheep, pigs, etc. are *not* capable of HOTs. But then why should we accept a HOT account of *phenomenal* consciousness? First, HOT theory gives a good explanation of the difference between conscious and non-conscious experience (so my experiences now are phenomenally conscious because available to or targeted by HOTs; whereas experiences in blindsight, absent-minded driving, sleep-walking, and during mild epileptic seizure are not so targeted). Second, HOT theory can provide a satisfying *explanation* of the phenomenal properties of conscious experience; which gives us reason to think that phenomenology, or 'feel', just *is* perceptual information available to, or targeted by, higher-order recognition.[5]

C1 is highly controversial, of course; and I make no pretence to have defended it here. It also conflicts with a powerful common-sense intuition to the contrary. But I suggest that this intuition may well be illusory, and can easily be explained away. For notice that one important strategy we often adopt when attributing mental states to a subject, is to try *imagining the world from the subject's point of view*, to see how things then seem. But when we do that, what we inevitably get are imaginings of *conscious* perceptions and thoughts, and of experiences with phenomenal feels to them. So, *of course* we naturally assume that the experiences of a cat will be *like* something, once we have got to the point of accepting (correctly, in my view) that the cat does have experiences. But this may merely reflect the fact that imaginings of perceptual states are always imaginings of *conscious* perceptual states, that is all. It may go no deeper than the fact that we have no idea how to imagine a non-conscious perception.

[3] See D. Rosenthal, 'Two concepts of consciousness', *Philosophical Studies* 49 (1986), pp. 329–59; and 'Thinking that one thinks', in M. Davies and G. Humphreys (eds.), *Consciousness* (Oxford: Blackwell, 1993), pp. 197–223; my *Language, thought and consciousness* and 'Natural theories of consciousness'; and W. Lycan, *Consciousness and experience* (Cambridge, MA: MIT Press, 1996). And see D. Dennett, *Consciousness explained* (London: Penguin Press, 1991), who also endorses a higher-order account, only with linguistic descriptions substituted in place of thoughts.

[4] See, e.g., papers in R. Byrne and A. Whiten (eds.), *Machiavellian intelligence* (Oxford: Oxford University Press, 1988); and D. Povinelli, 'Chimpanzee theory of mind?' in P. Carruthers and P. K. Smith (eds.), *Theories of theories of mind* (Cambridge: Cambridge University Press, 1996), pp. 293–329.

[5] See my *Language, thought and consciousness*, chapter. 7; and my *Phenomenal consciousness: a naturalistic theory* (Cambridge: Cambridge University Press, 2000).

Let me stress again that it will not matter very much, for the purposes of this paper, whether or not A1 and A2 are accepted, or thought plausible. My main interest is in the *conditional* question Q1; and to assess this you do not need to believe in, or accept the plausibility of, the antecedent. Another way of raising essentially the same conditional question, in fact (although without the same practical import), is to ask whether those inventions of the philosophical imagination, *zombies*, would be appropriate objects of sympathy and concern. (A zombie is someone who is supposed to be functionally and intentionally isomorphic to a normal person, but who is completely lacking in phenomenal consciousness, or 'qualia'.) Again, it does not matter whether or not zombies are really possible. They just provide another way of raising the general theoretical issue which is the concern of this paper: if we distinguish between the *feel* and the *functional role* of our sufferings and disappointments (which normally have both, of course), can we form a view about which (if not both) is the appropriate object of sympathy?

1.2. Assumptions about value

One route to a negative answer to Q1—which I want to close off with my third assumption—would be to endorse what Parfit calls an 'Objective List' theory of value,[6] and to maintain that the sufferings and disappointments of non-human animals figure on this list, irrespective of questions of phenomenology. So it might be claimed that pleasure and desire-satisfaction are moral goods, and pain and desire-frustration moral evils, quite apart from the question whether or not these states possess phenomenal properties, or 'feel'.[7] In which case we are morally obliged not to cause suffering to non-human animals, if we can help it, even if their mental states should prove to be lacking in phenomenology, as A1 and A2 entail.

The problem with an Objective List theory, however, is that it commits us to a form of metaphysical realism about value which is very hard to believe, as well as rendering ethical epistemology mysterious.[8] And there are a number of viable alternatives on the market. One would be to maintain, as Singer does,[9] that moral value is grounded in rationalised sympathy. Another would be to claim, following Scanlon,[10] that moral values and requirements are those which no one could reasonably reject who shared the aim of reaching free and unforced agreement; and this approach, too, will arguably find a central place for sympathy, embedded at the heart of the moral virtue of beneficence (to be contrasted with justice).[11] I therefore make the following assumption:

[6] See D. Parfit, *Reasons and persons* (Oxford: Oxford University Press, 1984), Appendix I.

[7] Of course there is a question whether a state which lacked 'feel' could properly be described as 'pain'—see S. Kripke, *Naming and necessity* (Oxford: Blackwell, 1980). But this semantic question is irrelevant to the main point at issue, which is whether a state with the functional and intentional/representational properties of pain, but without the distinctive phenomenology, could be an appropriate object of sympathy and concern.

[8] See Mackie, *Ethics*, Part One.

[9] See P. Singer, *Practical ethics* (Cambridge: Cambridge University Press, 1979; second edition, 1993).

[10] See T. Scanlon, 'Contractualism and utilitarianism' in A. Sen and B. Williams (eds.), *Utilitarianism and beyond* (Cambridge: Cambridge University Press, 1982), pp. 103–128.

[11] See my *Animals Issue*, chapter. 7.

A3: Metaphysical realism about moral values is false; rather, such values are somehow *constructed* from a basis in human attitudes, or human reason, or both.

The issues surrounding A3, and the lines of argument available in its support, are broadly familiar, if not universally accepted. I therefore feel justified in leaving A3 as just that—an *assumption* of the paper, without requiring anything further in the way of elaboration and comment.

It might be felt that there is another position which could accord moral standing to animals without becoming entangled in questions of phenomenology. This would focus on the idea of an *interest*, claiming that things which go against an animal's interests are appropriate objects of sympathy and moral concern, irrespective of issues to do with animal subjectivity. I believe that this position faces a dilemma: either it conceives of interests as being fixed, and as obtaining their value, independently of us—in which case it collapses into the sort of realism rejected in A3; or the moral significance of interests depends upon the mental states of the creatures in question—in which case we do not have a real competitor for the idea that *psychological harm* is the most basic object of sympathy and concern. This then gives me my fourth assumption:

A4: Attempts to ground sympathy in *interests* either collapse into a form of moral realism, or covertly appeal to the more basic importance of psychological harm.

This assumption raises large issues, to which I cannot hope to do justice within the scope of this paper. But I can say just a few words to motivate it.

Consider the claim that good health is in my best interests. What might this mean, and how might such a claim be grounded? It might, in the first place, be a claim about the intrinsic value of health, independent of my attitudes, feelings and goals, both now and in the future. This would be a very strong claim to make. And it is hard to see how it could be defended without commitment to the idea that *health* figures on some 'objective list' of goods. And the claim is, in any case, counter-intuitive in its own right. Of course it is generally true that poor health has a bad impact on the lives of people, even in cases where it is caused by the satisfaction of powerful desires, like desires for nicotine or for chocolate. But it is possible to conceive of cases where—unusually—someone's life-goals positively *require* poor health; in which case it is no longer true that good health is in their interests, it seems to me. A range of examples might be considered here. But imagine someone living in a country with a system of conscription of the healthy into the armed services, where that country is fighting a bitter and bloody (and, let us stipulate, unjust) war. Is it really in this person's best interests to be completely healthy in these circumstances, if that will mean, almost inevitably, severe psychological distress, together with a significant chance of painful injury and death?

It is true, of course, that interests are independent of *present* desires. It is familiar that health can be in my best interests even though I do not presently desire it. But in these cases, I think, the source of our concern derives from the belief that loss of good health will be seriously *regretted* in the future, or will otherwise have a negative impact upon my psychological life. That is to say, it derives from a belief in the psychological harm which poor health will bring in its train. But then this is no longer a competitor to the idea to be explored in this paper, that the proper object of sympathy is some sort of psychological

harm. And the same question arises, as to whether such harm, in order to be an appropriate object of sympathy, must involve some kind of subjective phenomenology.

II. Harm and the Frustration of Desire

In this section I shall defend two further assumptions, this time about the nature of psychological harm. The first is that the most basic form of such harm consists in the frustration of desire; the second is that only frustrations of desire which are *subjective* (in a sense to be explained) should count as harms.

I should stress that my focus throughout will be on present-tensed harm, rather than on the sort of harm which consists in longer-term psychological damage. For it is obvious that the former is the more fundamental phenomenon—if things could not be bad for me *at* a time, abstracting from any future effects on my life, then they could never be bad for me *over* time either; indeed, it seems plain that the most basic way in which an event can be bad for me in the longer term is by causing present-tensed harms at each of a number of later times. This is fully in accord with common-sense belief, I think. If I suffer intense pain for a short period of time, then everyone would allow that I am harmed psychologically, and that my state is an appropriate object of sympathy and preventative action, even if I am caused to forget about the occurrence of that pain immediately it finishes, and even if it has no further detrimental effects on my life.

I should also stress that a full account of the phenomenon of psychological harm is likely to be a complex business, the elucidation of which would take me well beyond the scope of this paper. To mention just one issue: we think that someone can be harmed, not just by being caused to suffer, but also if their life is caused to go less well, psychologically speaking, than it would have done otherwise.[12] So sometimes, at least, judgements of harm can involve cross-world comparisons of psychological goods and evils. But again it should be plain that the basic notion, on which I concentrate here, is the one which needs to be fed into such comparisons—namely, that of the goodness or badness of a psychological state for a subject *in* a given world.

II.1. Sensation versus frustration

First, let us consider what is the most basic form of psychological harm—is it the sensations characteristic of suffering (paradigmatically pain)? or is it, rather, frustrations or thwartings of desire? Two considerations demonstrate that the correct answer to this question is the latter, I believe. The first derives from the existence (or at least possible existence) of pure masochism. While most of us shun pain, the masochist welcomes it. In *im*pure forms of masochism, sensations of pain are welcomed, not for their own sake, but rather for their consequences, or because of their causally indispensable position in a desirable package of desires and other sensations (of sexual excitement, say). In such

[12] It is arguable that this idea can explain some of the intuitions which might otherwise seem to support the Objective List theory of value. For example, consider Rawls's man who just wants to count the blades of grass in various lawns, rather than become a mathematician (discussed in Parfit, *ibid.*)—although his only actual desire is satisfied, we might naturally think of that desire as itself standing in the way of a more worthwhile (in the sense of psychologically satisfying) life for the man.

cases we may say that the sensation of pain is itself unwanted, but is welcomed in context—somewhat as one might, in context, welcome the pain involved in the lancing of a boil: it is unwanted in itself, but it forms an indispensable part of a welcome event.

In cases of *pure* masochism, however, sensations of pain are themselves welcomed.[13] This need not mean that the pure masochist welcomes each and every pain, of course, irrespective of context and circumstances. A pure masochist need not be someone who enjoys toothache, or the pain of a gouty toe. Rather, it may only be pains surrounded by certain other sensations and satisfactions which are welcome. For example, it may be that only pains which are deliberately inflicted by another person in the context of sexual activity are welcomed. But in these cases, the sensations of pain are not just *tolerated*, as they are by the impure masochist, as being an indispensable part of a package which is worthwhile overall. Rather, they make a further addition to the satisfactoriness of the overall set of sensations and satisfactions. The pure masochist would not wish the sensations of pain away, even if everything else in their circumstances and sensations could remain the same.

I believe that there probably are some pure masochists in the world.[14] But I do not need to argue for this here. It is enough that pure masochism seems plainly conceivable, or conceptually possible. For what this then shows, is that it is not the *sensation* of pain (that is, its phenomenology, or felt quality) which is bad, in the normal case. It is rather that most of us want very much not to *have* that sensation. So it is the frustration of this desire which constitutes the harm, not the experience of the sensation itself.[15] This gives us the following claim:

> *A5:* Pains (and other sensations characteristic of suffering) only count as harms to the agent to the extent that they are unwanted by that agent; and the most basic form of psychological harm consists in frustrations or thwartings of desire.

This claim can be further defended in the light of the effects of certain types of analgesic, as Dennett once pointed out.[16] People under the influence of certain types of morphine report that the *sensation* of pain remains exactly the same; but that they no longer *care*. They say that the felt qualities of their pains remain just as they were before they were given the morphine, but that the presence of these qualities no longer bothers them. Such reactions are by no means arbitrary, inexplicable, or insane. On the contrary, they are predictable from, and explicable in terms of, the physiology of pain perception.

[13] Some might then wonder whether the state that the pure masochist welcomes is properly describable as 'pain'. Here, as before, the semantic issue is beside the point—which is that what the pure masochist welcomes has the characteristic phenomenology, of felt characteristics, of pain.

[14] See V. Warren, 'Explaining masochism', *Journal for the Theory of Social Behaviour* 15 (1985), pp. 103–129.

[15] What about the suggestion that it is really a frustration-sensation *pair* which constitutes the harm? This has not strictly been ruled out. But since in other cases—where the desire is not for the presence, or absence, of a sensation—frustration of desire alone can constitute a harm, it is simplest to suppose that this is so in the case of pain too. The question whether such frustration must itself have phenomenological properties is the main topic of the remainder of the paper.

[16] See D. Dennett, 'Why you can't make a computer that feels pain' in his *Brainstorms* (Hassocks: Harvester Press, 1978), pp. 190–229. This is the second of the considerations in support of A5 which I spoke of above.

As is now widely known, pain perception is mediated by two distinct nervous pathways—there is the *old path*, which is slow and projects primarily to the sub-cortical limbic system of the brain, and which seems to be responsible for pain *motivation*, underlying aversion to pain; and there is the *new path*, which is faster and which projects to a number of different cortical centres, and which seems to underlie fine discrimination, location, and feel.[17] What some types of morphine can do, is suppress the old path, while leaving the new path fully functional. So pain *perception*, in the sense of discrimination and feel, is unaffected, while pain *aversion* is suppressed. In such circumstances it seems plain that there is nothing bad, or harmful to the agent, in undergoing a mere pain sensation. Indeed, it really does seem appropriate to regard such types of morphine as an analgesic, removing pain in the sense in which pain *matters* to the agent.[18]

II.2. Objective versus subjective frustrations

While the two considerations presented above seem sufficient to establish A5, it should be noticed that A5 itself contains an ambiguity, depending upon how desire *frustration* is conceived of. For we can distinguish between *objective* and *subjective* frustrations of desire. The first occurs whenever the desired state of affairs fails to obtain or come about, *whether or not this is known of by the agent*. The second occurs whenever the agent *believes* that the desired state of affairs fails to obtain or come about, whether or not this is actually the case.

Everyone will allow, I think, that subjective frustrations of desire are a species of psychological harm. Everyone will grant that it is bad for agents, *ceteris paribus*, to think that things which they want have failed to occur. But some maintain that it is *also* a harm when an agent's desires are *objectively* frustrated, whether or not they are also subjectively thwarted.[19] If we held that objective frustrations of desire were a species of harm, and hence that such frustrations are worthy of sympathy and moral concern, then the answer to Q1 should obviously be negative. For then what makes frustrations of desire an appropriate object of concern would have nothing to do with phenomenology or feel. It would be the mere objective fact that an animal is in an unwanted state (e.g. pain) which is bad, irrespective of whether there is any phenomenology associated with the thwarting of its desire for the absence of the pain, and irrespective of whether or not the animal knows that it is in pain, or that its desire is being frustrated. Similarly, *mutatis mutandis*, for the

[17] See, for example, J. Young, *Philosophy and the brain* (Oxford: Oxford University Press, 1986).

[18] Note that the existence of a *new path* in many species of mammal is not immediately sufficient to establish that these species have pains which possess phenomenological properties. For it is one thing to *perceive* or to *discriminate* states of the world or of the organism's own body—which is what the presence of a new path strictly establishes—and it is quite another thing for these states of perception, or these discriminations, to be conscious ones—which is what is necessary for them to *feel like* anything, I maintain. According to the HOT theory of mental-state consciousness which lies behind A1 and A2, perceptual states only get to feel like anything when they are present to a conceptual system which is capable of classifying these states *as such*—as opposed to classifying the states of the world, or of the body, represented by those states.

[19] See T. Nagel, 'Death' in his *Mortal questions* (Cambridge: Cambridge University Press, 1979), pp. 1–10; and J. Feinberg, 'Harm and self-interest' in P. Hacker and J. Raz (eds.), *Law, morality and society* (Oxford: Oxford University Press, 1977), pp. 285–308.

case where an animal is starved of food or water—it would be the mere fact that its desires for these things are objectively thwarted which constitutes the existence of a psychological harm to the animal, irrespective of anything subjective.

However, I believe that objective frustrations of desire are *not* a species of harm. I propose, in fact, to defend the following claim:

> *A6:* Only subjective frustrations or thwartings of desire count as psychological harms, and are appropriate objects of sympathetic concern.

But this will require some setting up, beginning with a small detour through the falsity of philosophical hedonism.

The hedonist's focus is primarily on the satisfaction rather than the frustration of desire; but it carries obvious implications for the latter. For the hedonist claims that every desire aims at its own subjective satisfaction—normally interpreted to mean the *feelings* of satisfaction to be gained from achieving the overt object of the desire. For well-known reasons this is false.[20] Most desires aim at objective states of affairs—eating an apple, gaining a degree, or whatever—where knowledge that those states of affairs have been realised may characteristically *cause* certain feelings of satisfaction, but where those feelings were not what were desired in the first place. This then suggests that what is a benefit or a psychological good—from the perspective of the agent—about the satisfaction of desire is *achieving the goal of the desire*, not the subjective phenomenology of desire-satisfaction. Indeed, in the case of many desires (e.g. the desire that my children should grow up healthy, or the desire that my latest book should be well thought of) what I desire is the existence of a certain objective state of affairs *whether or not I ever come to learn of it.* So this then suggests that it is *objective* satisfactions of desire which constitute the primary form of psychological good. And then by parity of reasoning we might expect that what would be bad about the frustration of desire would be, not the phenomenology of frustration, or the knowledge that a desired state of affairs has failed to materialise, but rather the objective failure to achieve a goal.

This argument moves too swiftly, however. For it conflates the question of what is good or bad *from the perspective of the agent*—that is, in the light of the content of agent's desires and goals—with the question of what is good or bad *for* the agent, or a harm *to* the agent. The former notion of *good from the perspective of an agent* can be constructed—very roughly—by listing the contents of the agent's major goals (which may or may not mention the agent's own subjective states). But the latter notion is distinguished by its ties with beneficence and sympathy—to act beneficently towards someone is to do something which is good *for* them; and sympathy for someone's state is only appropriate for something which is bad *for* them, or a harm *to* them. So it would be perfectly consistent with the fact that my desires *aim at* objective affairs, to claim that the satisfaction of those desires should only count as a benefit, from the perspective of a beneficent observer, when I come to learn that the states of affairs in question have been

[20] See J. Feinberg, 'Psychological egoism' in his *Reasons and responsibility* (New York: Wadsworth, 1985), pp. 78–96, for example.

realised. Similarly, it would be consistent with allowing that it is the objective failure of my goals which is bad from my perspective, to claim nevertheless that the frustration of those desires should only count as a harm, from the perspective of a sympathetic observer, when their failure is known to me.[21] Not only would such claims be consistent, but they have a good deal of intuitive support. It is very hard indeed to see how anything could be good or bad *for* agents, or be a harm *to* them, without having any impact upon the (subjective psychology of) the agents themselves.

Admittedly, there are cases which can make it seem plausible that objectively frustrated desires are bad for—or constitute harm to—an agent, irrespective of their impact upon the subjectivity of the subject. Consider, for example, the case of the unknowingly cuckolded husband. Suppose that I love my wife, and want her to be faithful to me. Then what I want is that she should *be* faithful, not merely that I should continue to believe that she is faithful; and this is what constitutes the good, from my perspective. But now suppose that, unknown to me, she has an affair, which in no way affects our relationship. Some people have the intuition that I am harmed, and that I am, in the circumstances, an appropriate object of sympathy. But one possible confounding factor here is that what my wife does exposes me to *risk* of subjective harm (in addition to being *wrong*, involving breach of trust and evincing a failure of friendship), since her infidelity may one day be found out; and it may be that this itself is a harm.

In fact I doubt whether exposure to risk of harm is itself a harm. For consider the matter after the fact: when I finally die, having never found out my wife's infidelity, and having lived happily with her throughout, I think it would be a mistake to say that her actions had harmed me. For the risk never materialised. But even if risk of harm *is* itself a harm, it is plainly one which is parasitic upon the possibility of subjective harming. So a creature which is incapable of being subjectively harmed will be incapable of being exposed to risk of harm either. And then just the same questions concerning the moral status of animals will arise, if animal mental states are lacking in subjective phenomenology.

When we consider examples where factors such as wrong-doing and risk of harm are controlled for, I believe that any temptation to say that objective frustrations of desire should be counted as harms is dissipated. We can imagine a case, for example, where a woman has left Earth forever on a space-rocket, which lacks any means of radio communication (hence, what now happens on Earth can never be known to her). One of her deepest desires is that her late husband should be suitably honoured on Earth, and to this end she had commissioned a magnificent public statue of him to stand in a city square. But some months after her departure from Earth the statue is struck by lightning and destroyed. It seems plain that this event does not constitute any sort of harm to her—although it leads to the objective frustration of an important desire—and that

[21] The standpoint of a sympathetic observer (which is the standpoint of beneficence) should be distinguished from that of a friend or lover. To love someone is, in part and to a degree, to be prepared to enter into and adopt their goals as your own. So my friends may feel an impulse to assist in the realisation of my goals even if I will never know of the achievement of those goals, and so even if my failure to achieve those goals would not be a psychological harm to me.

sympathy for her situation is inappropriate. It seems equally plain that I do not act benevolently towards her if I see to it that the statue is rebuilt.[22]

Some readers may wonder whether these examples only *seem* to work because they are, in some sense, 'non-serious'; and may suggest that our intuitions will be different in cases where the objective frustration of a desire is of fundamental importance to the subject. To test this, imagine that the woman had left Earth on a space-rocket, not from choice (she is very distressed at having to leave), but because this is the only way she can earn enough money to pay for her children's education (she lives in an unenlightened land with no public education system; has no other realistic source of income; etc.—the example can be fleshed out). But then soon after her departure, her children are killed accidentally in a house-fire. Should we not feel sympathy for her, despite the fact that she will never know? Here, however, the confounding factor is that her children's deaths make her own sacrifice *unnecessary*; and it may be for this reason that our sympathy is engaged. This hypothesis can be confirmed. For we can switch the example so that she does not leave Earth *in order to* provide for her children, but for independent reasons, because she cannot stand to remain amongst the ungodly (or whatever). In such circumstances we no longer feel sympathy for her when her children are killed, I think, even though we know that she would have been devastated had she known.

I conclude, therefore, that only subjective frustrations of desire—frustrations which have a subjective impact upon their subject—count as harms, and are worthy of moral concern. This certainly makes it sound as if frustrations of desire which are lacking any phenomenology may therefore be *in*appropriate objects of concern. For what is it for a frustration to be *subjective* except that it possesses a phenomenology? In which case frustrations of desire which are non-conscious, and so which lack any phenomenology, will be *non-subjective* frustrations; and so, by A6, they will be inappropriate objects of concern.

II.3. Two kinds of subjective frustration

However, the sense of 'subjective' in A6 need not be—or not without further argument, at least—that of possessing phenomenological properties. Rather, the sense can be that of being *believed in by the subject*. On this account, a desire counts as being subjectively frustrated, in the relevant sense, if the subject *believes* that it has been frustrated, or *believes* that the desired state of affairs has not (and/or will not) come about. Then there would be nothing to stop a phenomenology-less frustration of desire from counting as subjective, and from constituting an appropriate object of moral concern. So we have a question:

> *Q2:* Which is the appropriate notion of *subjective* to render A6 true?—(a) possessing phenomenology? or (b) being believed in by the subject?

[22] Matters may be different if I am, not a stranger, but rather the woman's friend or lover. Here I might well feel an obligation to see the statue rebuilt. For as I noted above, to love someone involves a preparedness to take on their goals as your own. This explains, I think, the attitude characteristic of bereaved people towards the desires of the deceased. That we regard fulfilling the wishes of the dead as a way on honouring them, shows not that objective satisfactions of desire are a moral good, but rather that love may involve identification with the loved one's goals.

If the answer to Q2 is (a), then animal frustrations and pains, in lacking phenomenology by C1, will not be appropriate objects of sympathy or concern. This would then require us to answer Q1 in the affirmative, and animals would, necessarily, be beyond the moral pale. However, if the answer to Q2 is (b), then there will be nothing in C1 and A6 together to rule out the appropriateness of moral concern for animals; and we shall then have answered Q1 in the negative.[23]

It is important to see that desire-frustration can be characterised in a purely first-order way, without introducing into the account any higher-order belief concerning the existence of that desire. For it is primarily the absence of such higher-order beliefs in the case of non-human animals which constitutes the ground for denying that their mental states are conscious ones, as we saw in section I.1 above. So, suppose that an animal has a strong desire to eat, and that this desire is now activated; suppose, too, that the animal is aware that it is *not* now eating; then that seems sufficient for its desire to be subjectively frustrated, despite the fact that the animal may be incapable of higher-order belief.

In fact there is no more reason for insisting that desire-frustration requires awareness that one *has* that desire, than there is for claiming that *surprise* (in the sense of belief-violation) requires awareness that one has that belief.[24] In both cases the co-occurrence, in one and the same agent at one and the same time, of two activated first-order states with directly contradictory contents, is sufficient to account for the phenomenon. (In the case of surprise, what one has is an activated belief with the content *that P* combined with a perception, say, with the content *that not P*. In the case of desire-frustration, what one has is an active desire with the content *that P* combined with an activated belief with the content *that not P*.)

Let me emphasise that it is the *co-activation* of a first-order desire and a first-order belief with contradictory contents which is sufficient for subjective desire-frustration in the sense of Q2(b), not necessarily co-*consciousness* or anything involving higher-order thoughts about the creature's own states. Of course someone may desire *that P* and come to believe *not-P* without their desire being subjectively frustrated, if they never put the two things together. (For example, the desire may be dormant at the time when they acquire the belief; and the belief may remain dormant on the next occasion when the desire becomes active.) What is sufficient for subjective frustration is that a desire and a belief with directly contradictory contents should both be active together in the creature's practical reasoning system.

III. Sympathy and Subjectivity

How should Q2 be addressed? If we try to enter sympathetically into the mind of someone whose sufferings and frustrations are non-conscious ones, what we draw, of course, is a

[23] Which is not to say, of course, that such concern is necessarily *required* of us, either. That issue is moot, and needs to be debated in another forum. For my own view, see my *Animals Issue*, chapter 7.

[24] Of course, Davidson has famously maintained that belief requires the concept of belief, in part on the grounds that surprise presupposes an awareness, on the part of the subject, of what one had previously believed; see D. Davidson, 'Thought and talk' in S. Guttenplan (ed.), *Mind and language* (Oxford: Oxford University Press, 1975), pp. 7–24. For a brief critique, see my *Animals Issue*, chapter 6.

complete blank. We simply have no idea how to imagine, from the inside, a mental state which is non-conscious. Since to imagine undergoing a mental state is to imagine what that state *is like*; and since only conscious mental states are *like* anything, in the relevant sense; it follows that only conscious mental states can be imagined. This certainly makes it seem as if animal sufferings and disappointments, in being non-phenomenal by C1, cannot be appropriate objects of sympathy or concern. At any rate, this is what I argued in my *Animals Issue*, chapter 8.

But it may be that this conclusion is too hasty. It may be that *sympathy* and *imagination* can, and should, be pulled apart. That they *can* be pulled apart is obvious, I think. Sympathy can surely be grounded in a purely third-personal understanding of someone's situation and/or mental states; and certainly a *desire to help* need not be preceded by any sort of imaginative identification. To decide whether they *should* be pulled apart, what we need to ask is: what is bad about the frustration of a desire, from the perspective of a sympathetic observer?—the phenomenology of frustration? or the fact of coming to know that the desired state of affairs has failed to materialise? For the latter sort of frustration can certainly be undergone, in some form, by non-human animals. So putting A6 and Q2 together, in effect, we have the question:

> Q3: What is bad or harmful, from the point of view of a sympathetic observer, about the frustration or thwarting of desire?—(a) the phenomenology associated with desire-frustration? or (b) the fact of learning that the object of desire has not been achieved?

Another way of putting the point is this: we should concede that whenever we *enter sympathetically into* the frustrations and disappointments of another creature we always, and inevitably, imagine mental states with phenomenology. But it may be that this has more to do with imagination than with morals. It may be that imagination, as a conscious activity, can only represent from the inside mental states which are conscious, and so which possess phenomenological properties. But this may have nothing to do with what properly grounds sympathy. It may be, indeed, that what makes sympathy appropriate has nothing to do with phenomenology, just as answer Q3(b) envisages.

III.1. An (unsuccessful) argument from subjective importance

One argument in support of Q3(a)—that it is the phenomenology associated with desire-frustration which constitutes the harm to an agent—can be constructed as follows. Surely not *all* thwartings of desire constitute present-tensed psychological harms, and are appropriate objects of sympathy and moral concern. In particular, the thwartings of *trivial* desires, or mere whims, are not. But then how are these to be distinguished, except in terms of subjective phenomenology? It appears that our *important* desires are those whose frustration gives rise to more or less intense feelings of disappointment; whereas a trivial desire is one whose known frustration gives rise to no such feelings. In which case it is only those thwartings of desire which are accompanied by a certain characteristic phenomenology which are worthy of sympathy.

If it is only the thwartings of important (or, at any rate, non-trivial) desires which constitute a form of psychological harm; and if what marks out a desire as important (or,

at any rate, non-trivial) is something to do with the phenomenology associated with its frustration; then it will turn out that it is the phenomenology associated with desire-frustration which is psychologically harmful, just as answer Q3(a) envisages. And then it will follow, if C1 is correct, that the pains and frustrations of non-human animals will not be appropriate objects of sympathy and concern, because lacking in the relevant sort of importance.

It is surely false, however, that the only way to distinguish those desires which are important (rather than trivial) for an organism is in terms of some sort of subjective phenomenology. It may be that such phenomenology is used as evidence of importance in the human case; but it is doubtful whether it is constitutive of such importance. In the case of human beings, it is true, we often rely upon the extent of felt disappointment, or psychological devastation, in gauging the importance of a desire; but it is doubtful whether such feelings are what importance really consists in. At any rate, it seems easy enough to carve out a notion of *importance* which is not phenomenological, but for the application of which phenomenology might be evidential.

If desires can be ranked in terms of *strength*—where strength is defined, not phenomenologically, but rather in terms of which of its desires an agent would choose to satisfy first, *ceteris paribus*, or at what cost—then we can characterise those desires which are important as those which are stronger than most, or something of the sort. This would then give us a notion of *importance* of desire which would be applicable to non-human animals, whether or not such animals are subjects of phenomenology. And I think it is plausible that we humans use felt disappointment as evidence of importance, so defined. At any rate, unless this can be shown not to be the case, we lack any argument in support of Q3(a).

III.2. An (unsuccessful) argument from the case of Penelope

One attempt at answering Q3 has failed. We need to try a different tack. What we need to do, in fact, is to devise some thought-experiments to peel apart the respective contributions of known (or believed) failure of achievement, on the one hand, and phenomenological frustration, on the other, in our beliefs about psychological harm. Of course, many doubt the value of thought-experiments; and some maintain that a good theory is worth a thousand intuitions. But it is hard to see any other way forward in the present case. For the notion of sympathy which is our target is not itself a moral one (rather, it is something which *feeds into* moral theory, rather than being *constructed by* it); so we cannot resolve the question which concerns us through considerations of moral theory. And in all the real-world cases involving humans where our sympathies are aroused, the two distinct notions of 'subjective frustration' which are separated in Q3 are co-instantiated; whereas in real-world cases involving non-human animals it may be that our sympathies are driven by false beliefs (given C1) about animal phenomenology. We have no option, then, but to engage in thought-experiments—albeit experiments which are supplemented and constrained, wherever possible, by arguments.

What kind of thought-experiment do we require? One possibility would be to consider cases in which one and the same agent has both conscious and non-conscious desires, asking whether the frustration of the latter constitute any psychological harm *to him*. This

was the kind of line I took in my *Animals Issue*, chapter 8, where I imagined the case of Penelope, who is unusual in only ever having non-conscious pains in her legs.[25] These pains cause her to rub the offended part, and in severe cases to scream and cry out, but all without any *conscious* awareness of pain on her part. I suggested that these pains are not appropriate objects of sympathy and concern, largely because Penelope herself is not bothered by them—she may find it inconvenient when she cannot walk properly on a broken ankle (and we may feel sympathy for her inconvenience), but she is not consciously distressed by her pains, and does not seek our assistance (other than physical assistance) or sympathy. In similar spirit, we could imagine someone who only ever has non-conscious desires, in some domain, suggesting that it is doubtful whether we should feel sympathy when these desires are thwarted, since such thwartings are of no concern to the (conscious) agent.

I now maintain, however, that it is not really appropriate to consider thought-experiments of this type. This is because conscious subjects are apt only to identify with, and regard as their own, desires which are conscious. This is so for essentially the reason which makes it difficult for *us* to identify with those desires—there is no such thing as imagining, from the inside, what it is like to entertain such a desire; and subjects themselves can only even know that they *possess* such desires by self-interpretation, just as we know of them by other-interpretation. So from the perspective of the conscious agent, non-conscious desires will seem to be *outside of* themselves. Such subjects could, then, quite easily be mistaken in denying that the frustration of a non-conscious desire constitutes any harm to them. In fact, they are in no better position to express a judgement on this matter than we are.

How, then, are we to peel apart thwarted agency from phenomenology in our beliefs about psychological harm? The way forward, I believe, is to consider, not examples in which one and the same agent has both conscious and non-conscious desires; but rather an example in which an agent has conscious desires, but where satisfactions and frustrations of those desires are lacking in any of the usual phenomenology. This is the example of Phenumb.

III.3. The case of Phenumb

Let us imagine, then, an example of a conscious, language-using, agent—I call him 'Phenumb'—who is unusual only in that satisfactions and frustrations of his conscious desires take place without the normal sorts of distinctive phenomenology.[26] So when he achieves a goal he does not experience any warm glow of success, or any feelings of satisfaction. And when he believes that he has failed to achieve a goal, he does not experience any pangs of regret or feelings of depression. Nevertheless, Phenumb has the

[25] Of course, we can again raise the (semantic) question whether what she has is properly described as 'pain'. But, as before, this is not to the point. Our question is whether a state which is at least *like* pain in respect of its functional role and intentional content, but which lacks any surroundings of subjective phenomenology, is an appropriate object of concern.

[26] Science fiction fans might identify Phenumb with Mr Spock from the television series *Star Trek*; or perhaps better, with some pure-blooded Vulcan.

full range of attitudes characteristic of conscious desire-achievement and desire-frustration. So when Phenumb achieves a goal he often comes to have the conscious belief that his desire has been satisfied, and he knows that the desire itself has been extinguished; moreover, he often believes (and asserts) that it was worthwhile for him to attempt to achieve that goal, and that the goal was a valuable one to have obtained. Similarly, when Phenumb *fails* to achieve a goal he often comes to believe that his desire has been frustrated, while he knows that the desire itself continues to exist (now in the form of a wish); and he often believes (and asserts) that it would have been worthwhile to achieve that goal, and that something valuable to him has now failed to come about.

Notice that Phenumb is not (or need not be) a zombie. That is, he need not be entirely lacking in phenomenal consciousness. On the contrary, his visual, auditory, and other experiences can have just the same phenomenological richness as our own; and his pains, too, can have felt qualities. What he *lacks* are just the phenomenal feelings associated with the satisfaction and frustration of desire. Perhaps this is because he is unable to perceive the effects of changed adrenaline levels on his nervous system, or something of the sort.

Is Phenumb an appropriate object of moral concern? I think it is obvious that he is. While it may be hard to imagine what it is *like* to be Phenumb, we have no difficulty identifying his goals and values, or in determining which of his projects are most important to him—after all, we can ask him! When Phenumb has been struggling to achieve a goal and fails, it seems appropriate to feel sympathy: not for what he now *feels*—since by hypothesis he feels nothing, or nothing relevant to sympathy—but rather for the intentional state which he now occupies, of dissatisfied desire. Similarly, when Phenumb is engaged in some project which he cannot complete alone, and begs our help, it seems appropriate that we should feel some impulse to assist him: not in order that he might experience any feeling of satisfaction—for we know by hypothesis that he will feel none—but simply that he might achieve a goal which is of importance to him. What the example reveals is that the psychological harmfulness of desire-frustration has nothing (or not much—see the next paragraph) to do with phenomenology, and everything (or almost everything) to do with thwarted agency.

The qualifications just expressed are necessary, because feelings of satisfaction are themselves often welcomed, and feelings of dissatisfaction are themselves usually unwanted. Since the feelings associated with desire-frustration are themselves usually unpleasant, there will, so to speak, be *more* desire-frustration taking place in a normal person than in Phenumb in any given case. For the normal person will have had frustrated *both* their world-directed desire *and* their desire for the absence of unpleasant feelings of dissatisfaction. But it remains true that the most basic, most fundamental, way in which desire-frustration is bad for, or harmful to, the agent has nothing to do with phenomenology.

III.4. Does consciousness make a difference?

If the example is accepted, then the case of Phenumb is successful in showing that the thwarting of at least a *conscious* desire can be bad for an agent in the absence of any phenomenology of frustration, as seen from the perspective of a sympathetic observer. We now need to ask whether this subjective badness has anything to do with the fact that the desire is a conscious one. Or would the frustration of a non-conscious desire, too—such as

non-human animals only have, by A1—be equally bad for the agent? In fact it is hard to see why consciousness, here, should make any difference.[27]

There are just two things which distinguish Phenumb's case from that of an animal, in fact (on the assumption that A1 and A2 are true). The first is that Phenumb has higher-order *beliefs* about (or knowledge of) the existence of his first-order desires, and their frustration; whereas by hypothesis an animal does not. The second is that Phenumb has higher-order *preferences* between, and evaluations of the significance of, his first-order desires; whereas again an animal will not. So we need to consider whether either of these differences can be used to block the inference from the fact that sympathy is appropriate for Phenumb, to the conclusion that it can be appropriate for the desire-frustrations of an animal. (And recall from section II.3 above, that desire-frustration can be characterised in an entirely first-order way.)

It is very hard to see how it could be the presence of a higher-order belief, in the case of Phenumb, which makes sympathy for his desire-frustrations appropriate. What is bad for Phenumb, surely, is that his desire is co-active with the knowledge that the object of his desire has not been achieved. It does not seem relevant that he knows that this was his goal—i.e. that he has a higher-order belief about his own state of desire. For what gets frustrated is the first-order desire. Of course, in the normal case, the presence of the higher-order belief may be sufficient to cause the feelings of disappointment which are normally consequent on the frustration of a desire. But we have already shown that it is not these which are relevant. Once we have taken the phenomenology of desire-frustration out of the picture, in fact, it becomes plain that the first-person badness of desire-frustration has nothing to do with the fact that the agent *believes* that he has that desire, and so comes to believe that a desire of his has been frustrated. The badness consists in the frustration, not the higher-order belief.[28]

It is more plausible that it might be Phenumb's possession of second-order desires or preferences which makes the relevant difference. For this at least is something conative as opposed to cognitive. But consider an ordinary case where someone's first-order and second-order desires are in conflict. Suppose that Mary is trying to give up smoking and wants very much (first-order) to have a cigarette right now while also wishing (second-order) that this desire should not be satisfied. It is surely appropriate that one might feel sympathy for Mary's unsatisfied first-order craving. But we have already established that what makes this sympathy appropriate cannot be the *phenomenology* of first-order desire-frustration. And in this case it plainly cannot be the frustration of any second-order desire which might warrant our sympathy, because Mary has no such desire. So all that is left is

[27] Note that consciousness *may* make all the difference if the question is whether sympathy is morally *demanded* of us. For it may be that consciousness is a necessary condition of full moral personhood, and that only moral persons (and those of the same species as moral persons) can morally *command* our sympathy in and of their own right—see my *Animals Issue*, chs. 5–7. This is not to the point here. The issue in the text is whether non-conscious frustrations of desire count as psychological harms *from the perspective of a sympathetic observer*. So what is at issue is, at most, whether such frustrations are *possible* objects of sympathy, not whether sympathy for them is morally *required* of us.

[28] Of course it might be maintained that what is relevant for sympathy is a higher-order belief and first-order frustrated-desire *pair*. But it is very hard to see what could motivate such a view.

that it is the bare first-order frustration which forms the object of our sympathy. And then we get a conclusion which will transfer to the case of animals.

IV. Conclusions

If my assumptions A3 and A4 are granted, then the main point is (at least tentatively) established: the most basic form of psychological harm, from the perspective of a sympathetic observer, consists in the known or believed frustration of first-order desires (which need not require that agents have knowledge that they *have* those desires—just knowledge of what states of affairs have come about). That is to say, the answer to Q3 is (b). So the proper object of sympathy, when we sympathise with what has happened to an agent, is the known (or believed) frustration of first-order desire. And it follows, then (given A1 and A2), that the non-conscious desires of non-human animals are at least *possible*, or *appropriate*, objects of moral sympathy and concern. (Whether they *should* then be objects of such concern is a further—distinctively moral—question, to be answered by considerations pertaining to ethical theory rather than to philosophical psychology.)[29] And it emerges that the complete absence of phenomenology from the lives of most non-human animals, derived in C1, is of little or no direct relevance to ethics.

What emerges from the discussions of this paper is that we may easily fall prey to a cognitive illusion when considering the question of the harmfulness to an agent of non-conscious frustrations of desire. In fact, it is essentially the *same* cognitive illusion which makes it difficult for people to accept an account of mental-state consciousness which withholds conscious mental states from non-human animals. In both cases the illusion arises because we cannot consciously imagine a mental state which is *un*conscious and lacking any phenomenology. When we imagine the mental states of non-human animals we are necessarily led to imagine states which are phenomenological; this leads us to assert (falsely, if C1 is true) that if non-human animals have any mental states at all (as they surely do), then their mental states must be phenomenological ones. In the same way, when we try to allow the thought of non-phenomenological frustrations of desire to engage our sympathy we initially fail, precisely because any state which we can imagine, to form the content of the sympathy, is necessarily phenomenological; this leads us (again falsely, if the arguments of this paper have been sound), to assert that if non-human animals *do* have only non-conscious mental states, then their states must be lacking in moral significance.

In both cases what goes wrong is that we mistake what is an essential feature of (conscious) *imagination* for something else—an essential feature of its *objects*, in the one case (hence claiming that animal mental states must be phenomenological); or for a necessary condition of the appropriateness of activities which normally *employ* imagination, in the other case (hence claiming that sympathy for non-conscious frustrations is necessarily *in*appropriate). Once these illusions have been eradicated, we

[29] For my own hedged-about-negative (or heavily qualified positive) answer, see my *Animals Issue*, chapter 7.

see that there is nothing to stand in the way of the belief that the mental states of non-human animals are non-conscious ones, lacking in phenomenology. And we see that this conclusion is perfectly consistent with according full moral standing to the sufferings and disappointments of non-human animals.[30]

University of Sheffield

Received: July 1998
Revised: March 1999

[30] An earlier version of this paper was delivered at the conference, *Perspectives on animal consciousness*, held in Wageningen, Netherlands, in July 1997. I am grateful to Marc Bracke and Robert Heeger, who were my respondents, and to all those who participated in the subsequent conference discussions. I am also grateful to David Archard, Keith Frankish, Christopher Hookway, Richard Joyce, Simon Kirchin, Robert Stern, Jo Wolff, and especially Leif Wenar for their helpful comments on earlier drafts; and to the referees for this journal whose perceptive comments on the submitted paper occasioned substantial revisions.