Ben Davies – Unfinished draft, please do not cite without permission

**1. Introduction**

Subjectivist theories of welfare claim that individual welfare depends in some way on mental attitudes such as desires or preferences. A simple desire theory (e.g. Brandt, 1972; Murphy, 1999; Heathwood, 2005) says that it is good for us to get what we want, and bad for us not to get what we want; a more complex desire theory might say that it is good for us to get what we want, but only given certain idealized conditions such as consistency or sufficient information about what getting what we want will actually involve, or on the stipulation that the desires are not brought about by alien influences such as mental manipulation or threats.

This paper considers two criticisms that, i) are aimed not just at one subjectivist theory, but subjectivism comprehensively, and ii) together pull the subjectivist in different directions with regard to an appropriate response. The first criticism, (Skelton, 2015; Lin, 2016) argues that subjectivist theories fail to be persuasive because they cannot plausibly be applied to young infants. To this criticism, we might well add that many subjectivist theories appear not to apply to non-human animals, or to adult humans who lack various mental capacities. Yet these groups are surely capable of achieving or failing to achieve personal welfare. So, we should reject subjectivist theories. Call this the Infant Problem.

The second criticism is sometimes called the ‘paradox of desire’ (Bradley, 2006; Bradley, 2009; Skow, 2009). This paradox suggests that under certain unusual but plausible conditions, subjectivist theories will be unable to coherently say whether someone’s lifetime welfare is positive or negative. A theory that fails to give a coherent description of a possible state of affairs should be rejected. So, we should reject all subjectivist theories. Call this the Paradox Problem.

In Sections 3 and 4, I outline the Paradox Problem, and explain why it should lead us to endorse a complex attitude as grounding welfare if we are to rescue subjectivism. As I then suggest, this strategy places it in direct tension with the issues raised by the Infant Problem. Subjectivists who feels the force of the Infant Problem should feel some pressure to have their theory rely on *simple* mental attitudes, the kind of which an infant is capable. So, these two problems pull subjectivists in apparently contradictory directions.

My suggestion, in Section 5, is that subjectivists should adopt a disjunctive theory, where welfare for infants (and some others) depends on different attitudes than welfare for cognitively developed adult humans. Section 5 also explains why a disjunctive theory is not merely an *ad hoc* response to the two problems.

Before doing that, however, it seems worthwhile to consider why we should want to preserve a subjectivist theory in any case. Why, in other words, should our reaction to these paradoxes not simply be to give up on subjectivism and embrace objectivism? Section 2 offers an incomplete response to that question.

**2. Why subjectivism?**

This section outlines several reasons for preferring a subjectivist[[1]](#endnote-1) theory of welfare. This is not a comprehensive defense of subjectivism. But since some of the theories I will consider will be complex, it is worth outlining some motivation for thinking that such theories might be right. I will also not be offering a comprehensive outline of the debate between objectivists and subjectivists. My outline will thus be fairly one-sided.

The first reason that attracts many people to subjectivism is that at least some of the alternatives – specifically, objectivist views that say that value properties are *irreducibly normative –* are metaphysically and epistemologically unattractive. As Mackie (1977: 38-42) says, the property of being good – really, objectively good – seems to be unlike anything else we confront in the natural world. Relatedly, it is not clear how we can come to know about normative facts; we cannot come to know about them using the methods with which we come to know ordinary naturalistic truths, and do not seem to have any reliable ‘moral sense’ with which to detect moral truths. Of course, suggesting a novel kind of phenomenon is not in itself a reason to throw out a theory. But if we can explain our judgements of objective value by appealing to subjective attitudes (e.g. Nichols, 2004; Joyce, 2006; Street, 2006), it seems reasonable to prefer an explanation that does not involve such odd properties (see also Garner, 1990; Horgan and Timmons, 1992; Joyce, 2001; Olson 2014).

### A second concern about objectivist theories is a worry about evaluative arrogance (see e.g. Railton, 2004). Consider Mill’s (1863/2003, Chapter 4) supposed proof of hedonism, which I take to be a kind of objectivism.[[2]](#endnote-2) Mill observes that the only evidence we could possibly have for thinking that something is desirable is that people do in fact desire it. Moore (1903, §39-41) famously objected to this idea that in Mill’s terms, desirable seems to mean merely ‘capable of being desired’ not, as it should, ‘worthy of being desired’. Defenders of Mill (Skorupski 1999, 220-3) have responded that this misunderstands his argument. For what evidence could we possibly have that something is worthy of being desired, other than that it is in fact desired? If we point to something nobody values, and which has no hidden connection to anything that people value, and claim that it is nonetheless something that people *should* value, we are guilty of a kind of evaluative arrogance.

But there is still a problem with this argument. Mill may be right that we have no *better* evidence than what people do in fact value. But this does not mean that this is good evidence. The fact that people value a certain thing is evidence only for the fact that they value it. In fact, this concern applies to all objectivist views that do not fall into the aforementioned category of those that classify value as irreducibly normative. The alternative family of views, that value is in some sense dependent on or reducible to some natural property, will always face the general problem of someone who does not value that property. For instance, some views attempt to derive objective evaluative facts from an account of human nature (e.g. Finnis 1980; Foot 2001); certain things are good because, say, they are naturally good for humans, or reflect natural human ends. Such views must explain why someone who does not enjoy, say, companionship should be convinced by the argument that *most humans* or *normal humans* do enjoy it, and so it must be good for them.

Even if Mill is right in his claim that we all value pleasure, we might still think it coherent to imagine someone who *did not* desire, value or take some other relevant attitude toward pleasure. At this point, Mill has two options. He can insist that pleasure is nonetheless good for this individual, at the cost of his proof. Or he can admit that pleasure is *not* good for this individual, staying true to his argument. But this would highlight that what drives the force of the argument is not some quality pleasure is supposed to have, but instead a quality of our relation to it, i.e. that we take some positive attitude towards it.

One might object that there is something incoherent about all this. For I have just claimed that we can coherently imagine a person who does not value pleasure. And one might think that this is a contradiction, since if a sensation is not valued to *some* degree, then it does not count as pleasure. What is at stake here is a conceptual issue about the nature of pleasure and pain. I am personally sceptical of the view that a sensation cannot count as pleasurable if it is not valued at all, or as pain if not at all disvalued. As Hall (1989) notes, there seems to be at least some evidence that certain pain ‘killers’ such as morphine do not actually remove any of the sensational components of pain; what they do is make it such that those sensations no longer bother the patient. What is unpleasant about pain, then, is that we take some negative psychological attitude towards it.

On the other hand, suppose that it is a conceptual truth that if we do not have some positive psychological reaction to a sensation, it cannot count as pleasure (and, *mutatis mutandis*, pain). While this puts paid to my challenge against Mill, since the idea of someone who did not value pleasure would be incoherent, it does not undermine the idea that what is good *about* pleasure is the fact that we take some positive attitude toward it. All that would be true on this view is that if subjectivism about the good is true, then pleasure is necessarily good for us. But it certainly does not show that it is objectively good for us.

To reiterate, I do not suppose that I have given anything like a comprehensive argument in favour of subjectivism about welfare, or value in general. Rather, my aim in this section is simply to outline some considerations that independently motivate the aim of finding a solution to the two problems that I will consider. In particular, this motivation is to find a particular kind of solution: one that gives no place whatsoever to objective value. Those who are not convinced by any of the issues I have just outlined will find this form of solution less pressing; they may be satisfied with simply giving some place to subjectivity in determining welfare (though as we will see, both problems present challenges even for this weaker position). My aim is to explore a solution to these problems for the thoroughgoing subjectivist.

**3. The Paradox Problem and the Infant Problem**

*3.1 The Paradox of Desire*

To outline the Paradox Problem, it is helpful to make a few assumptions. While the resulting theory will be simple, this should not be taken to imply that more complex subjectivist theories do not also fall prey to the paradox; the simplicity of the theory-form I discuss is for explanatory purposes.[[3]](#endnote-3)

First, we must stipulate how a person’s attitudes contribute to her time-relativewelfare. By time-relative welfare, I mean how well off or badly off a subject S is at a particular time *t*. For the desire theory, assume that a satisfied desire contributes to, and a frustrated desire detracts from, welfare in line with the intensity of that desire, at the moment the desire is satisfied or frustrated. If S desires an apple with intensity that we can (arbitrarily) represent with a strength of five,[[4]](#endnote-4) then getting the apple will give S five welfare points and failing to get the apple will cost S five welfare points.

The next assumption concerns how to assess the overall balanceof temporal welfare at any one moment, since several desires might be satisfied and/or frustrated at the same time. At any one moment, assume that S must have one – and only one – level of welfare, which must be positive, negative or neutral, the latter of which we can represent as zero. If S has multiple desires at any one moment, some frustrated and others satisfied, S’s welfare at that moment is the balance of these values, determined by their addition. If S has two desires, one with an intensity of ten that is satisfied, and one with intensity five that is frustrated, she has an overall positive welfare of five.

A third assumption is that time-relative welfare values contribute to S’s ‘extended’ welfare in an additive way. S’s extended welfare is her total welfare across times *t1…tn*., which includes an assessment of how well her life goes overall. Lifetime welfare is thus the sum of time-relative values across an entire life. For instance, if S’s life consists of three moments, whose values (given the balance of various frustrated and satisfied desires at those moments) are five, minus ten, and five again, her lifetime welfare is zero, and her life was neither good nor bad overall.

Here is an example, adapted from Bradley (2009: 30-46), of how desire theories lead to the paradox, expressed in a way that makes the above assumptions:

S has some ordinary desires over the course of her lifetime, whose satisfaction and frustration happen to have balanced out exactly. According to the desire theory, her welfare is currently zero. She forms two further desires. One is for a nearby glass of water (intensity 5). The other is that her life should go badly overall (intensity 10). She reaches out, spills the water, and dies.

What was S’s lifetime welfare? She starts the final moment of her life at the neutral level. A desire is frustrated, which by itself would bring her welfare to -5. Her other desire is that her life should go badly overall; since she is currently at a level of negative welfare then her desire should, if nothing else changes, be satisfied. The paradox emerges because if her life does indeed go badly overall, and that desire is satisfied, then she gains welfare of 10; according to the desire theory, this means that her lifetime welfare moves to 5. But this entails that her life goes *well* overall, frustrating the original desire, which means her lifetime welfare is -5, which means that her life has in fact gone badly overall. And so on. Desire theories seem unable to give a single, coherent answer about whether S’s life has gone well.

Bradley claims that this result applies to all subjective theories of welfare. The same conclusion seems to follow if S has a preference that her life goes badly overall; sets it as a goal; wishes for it; or hopes for it, so long as we have the satisfaction and frustration of attitudes contributing to welfare as outlined. What is more, this paradox is not only problematic for theories that base welfare solely onattitudes. A pluralistic theory that, for instance, combined hedonism with desire satisfaction could meet the same problem for an agent whose lifetime balance of pain and pleasure was exactly equal. So, the paradox seems to suggest that we should exclude subjective attitudes *completely* as direct determinants of welfare.

My suggestion is that a particular form of subjectivism, outlined by Dale Dorsey (2012), is the most promising subjectivist response to the Paradox Problem, though it will need some development. Bradley (2007: 47) claims that the theories for which the paradox is problematic are united by a common feature. This is that according to each, a person’s welfare is determined by “(i) the person’s attitudes towards states of affairs or propositions...and (ii) whether those states of affairs are *true*”. This is a problem because S’s attitude to her lifetime welfare affects its own truth status in a way that jars with its role as a constituent of welfare. The thought is that the insistence on truth or correspondence generates a problem, and will do so for all subjectivist theories.

However, Dorsey’s judgment subjectivism meets Bradley’s criteria for a subjective theory, but does not seem to generate the Paradox Problem. Judgment subjectivism says that S’s welfare is increased when she judges something to be good for her, and that thing occurs (Dorsey 2007: 47). But judging something to be good for you is notan attitude one can coherently hold towards one’s own life going badly. S might desire that her life go badly without believing that occurrence to be good at all; perhaps she has an uncontrollable urge for things to go badly for her. Alternatively, she might think that her life going badly is good in an impersonal sense (maybe because she is a bad person, and it is right that bad people have bad lives), or that her life going badly will, somehow, be good for some other individual or cause. But while it is thus coherent for S to desire (hope; wish; set as a goal) that her life goes badly, and even to judge it good in some sense, it is notcoherent for S to judge it *good* *for her* that her life goes badly. As Dorsey (op cit: 423) puts it, judgment subjectivism says that “for ɸ to make me worse-off, I *must* believe that it will make me worse-off, rather than better-off. Because any set of coherent evaluative beliefs[[5]](#endnote-5) will rule out the inconsistent set ‘ɸ is intrinsically better for me’ and ‘ɸ is intrinsically worse for me’, judgment subjectivism does not allow that my welfare can be increased by its decrease.” On this view, then, if S claims to judge that her life going badly is good for her, she is making a conceptual mistake. S can certainly *desire* that her life go badly overall; but it must be for some reason other than that she judges that it would be good for her. None of the judgments outlined above that might make sense of S’s desire for her life to go badly involves S taking it as good *for her* that her life goes badly.

As such, the paradox cannot get started. For while S may desire both the water and that her life goes badly, only one of those is relevant to her welfare according to judgment subjectivism. Since she fails to get the glass of water, which we can stipulate she judged to be good for her, her life goes badly overall. So, her second desire is satisfied. But since this second desire did not involve a coherent judgment that its satisfaction would be good *for her*, there is no further effect on welfare, and so no paradox is generated. Judgment subjectivism can give a coherent answer to the question of S’s lifetime welfare in a way that other subjectivist theories cannot: her life went badly overall.

However, it should be clear that judgement subjectivism is a sitting duck for the Infant Problem. As Lin (2017, 357) says, “Newborns don’t have welfare beliefs...Although recent advances in developmental psychology have shown that older babies are surprisingly sophisticated, there is no evidence that *newborn* babies have welfare beliefs. They like some things and are averse to others, and perhaps they have beliefs. But they surely do not have beliefs to the effect that X is good for them.” If that is right, then judgement subjectivism avoids the Paradox Problem only at the cost of walking directly, and obviously, into the Infant Problem.

In fact, things are even worse for judgement subjectivism than Lin suggests. For as I will now argue, the judgement that Dorsey describes is not sufficient to avoid a reformulated version of the Paradox Problem. A response to this version seems to require an even more cognitively complex attitude, which is even less likely to be attainable by newborns.

The problem is as follows. Even if it is incoherent to judge it good for you that your entire life goes badly, it may not be so obviously incoherent to judge it good for you that a shorter *period* of your life goes badly, so long as that badness contributes to greater good later. The Paradox Problem can thus arise across times shorter than lifetimes, such that the response to the problem outlined above cannot work. There is a solution to this – restricting the relevant judgments to judgments of *intrinsic* value – but this faces additional pressure from the Infant Problem.

It will be quite rare to find a genuine example of the relevant kind of judgement. It is not enough that it is good for S to take some course of action that, although it will have a bad effect at the time, promises even greater good later on. Examples of this kind of trade-off are commonplace; but they do not involve the judgment that the period of time *going badly* is itself what is good for S. For instance, if I judge that quitting smoking cold turkey will be unpleasant at first, but sufficiently better for me in the long run, it is not the initial badness that I judge good for me. If I could quit cold turkey without that being unpleasant, that would not change my judgment of the good effects at all.

To generate this variation of the paradox, then, we must turn once again to somewhat artificial formulations. Here is an example. I promise S that I will give her a thousand pounds if the next minute goes badly for her. S judges that it will be good (with a value of ten) for her if her life goes badly in the next minute. In that minute, she makes just one other judgment – that it would be good (with a value of five) for her to drink the glass of water on the table. She spills the water just before the minute ends. Has the minute gone well or badly for her? It seems that if it has gone badly, then her judgment that it would be good if the minute went badly has been realized. This means that the minute went well because she will now get her prize money. But if the minute went well, then it didn’t go badly; this means that the state of affairs she judged to be so good for her – the minute going badly for her – didn’t occur. So her judgment is not realised and she will not get her money. But then the minute went badly because all it contained was a failed attempt to get a glass of water, and so on. The paradox resurfaces.

S’s judgments about the case make a kind of sense in a way that her judgment about her life going badly does not seem to. Since something quite considerable is riding on that minute going badly for her, it surely *is* good for her when the minute goes badly for her. So, it is not clear that we can appeal to any inherent contradiction in her judgments, as I did in the lifetime case. Unlike the case of quitting smoking, it does seem plausible to say that it is the minute’s going badly *itself* that S judges good for her. Whereas I could get exactly the same good results from quitting smoking if it was not unpleasant, the case above is set up so that S really can coherently judge that it is good that the minute has gone badly for her. This really does seem to be a case where some period of time going badly for an individual is good for her.

However, the fact that S seems able to coherently judge that the minute going badly for her was good for her should make us look again at the case. This judgment only seems to become coherent once we introduce the idea of a motivation external to the period of time being judged. That is, it does seem incoherent for S to judge that it is good for her *at that minute* that the minute goes badly for her. For her to coherently judge that the minute going badly is good for her, it seems that it must be good for her because of something it delivers at some other time.

This suggests a response somewhat analogous to something that Bradley says about welfare hedonism. Bradley (2009: 20) argues that ɸ is bad for S at all times when, and to the extent that, it causes her pain or prevents her from experiencing pleasure. He considers the objection that some things can be bad for you without either of these being the case. Drinking poison yesterday, for example, might cause S great pain today. Isn’t there a sense in which drinking the poison was bad for S, even though it didn’t *then* cause her pain? Bradley insists plausibly that although the poison caused today’s pain, it is not *constitutive* of the pain’s badness. That pain would be just as bad if it had been caused by food poisoning, for example. To say that drinking poison was bad independently of the pain is to count it twice: once when it occurs and once when it has its bad effect. Drinking the poison is bad for S in a *derivative* sense; but we can coherently deny that it affects her welfare since it is not itself a bad experience, and hedonism claims that welfare is purely experiential.

It seems that judgment subjectivists can say something analogous about S in the case of the shorter Paradox Problem. Note that unlike the desire theory, the apparent incoherence in the judgment version of the paradox seems to occur at the level of the agent’s judgments, not at the level of the theory. After all, S *only* judges that the minute was good for her because she *also* judges that it was bad for her. So when she judges that the minute was good for her, she does not *stop* judging that it was bad for her. If she did, she would also have to abandon her judgment that it was good for her. Either S is making an incoherent judgment overall – and this does not seem to be the case – or there are in fact two distinct kinds of judgment at play. If I confront S about her judgment that the minute went well, perhaps in attempt to avoid paying up, she might well say respond, “Yes, the minute went well for me *in a sense*, because it won me your money.But it went badly in the sense we meant when we made the agreement”.

Judgment subjectivists can appeal to this apparent distinction between two ways of judging ɸ to be good for you, and only base welfare on one kind. That distinction seems analogous to Bradley’s distinction between causal and constitutive value – while S’s minute causessomething that she judges good for her, it is not constitutively something that she judges good for her. What S judges to be good for her is the receipt of a large sum of money. The minute going badly is only good derivatively. And just as in Bradley’s poison example, claiming that such derivative value intrinsically enhances your welfare is to count certain sources of value twice. The receipt of the money gets counted once when S receives it, and once when it determines the value of the minute going badly (after all, there is no *other* sense in which that minute is good for her, because only one thing happened during the minute, and it was bad for S). As such, judgment subjectivism can avoid the Paradox Problem in cases where it does not seem strictly incoherent to judge that something’s being bad for you is good for you by claiming that only judgments of *intrinsic* or *constitutive* value are relevant to welfare.

*3.2 The Paradox of Judgement?*

Judgement subjectivism is thus able to avoid the paradox of desire. However, since the paradox of desire is generated by (indirectly) making use of the *second-order attitudes* (i.e. desires that turn about to be about desires, at least according to the relevant theory of welfare), we might wonder whether we could not also generate an analogous problem for judgement subjectivism, since this is also a theory that depends on attitudes.

Such a view would depend on the idea of second-order judgements of the good; judgements, that is, about how your judgements of what is good for you are themselves good for you. And it seems clear that we can generate such a judgement that has at least the appearance of paradox. Consider the following situation:

P makes a number of judgements about what is good for her. Over the course of her life, exactly 1,000 of these judgements are realised, while the rest are not. The final judgement she makes is that it would be good for her if *an even number* of her judgements about what is good for her are realised. Call this final judgement, G.

It is hopefully clear why we have an apparent paradox. Given the facts about her previous judgements, G appears to be realised. But this means that she has a further judgement about what is good for her realised, taking her total up to 1,001. That in turn means that an odd number of judgements have been realised, making G *not* realised, bringing the number back down to 1,000. And so on. All that is required to make an equivalent paradox for judgement subjectivism is that – G aside – P’s lifetime welfare is neutral. Now judgement subjectivism cannot say whether the life of a person with conceivable (albeit even odder than S’s) attitudes is overall good, bad or neutral.

The first thing to note, however, is that this paradox comes about in a different way to the paradox of desire. The paradox of desire, at least as Bradley describes it, comes about *indirectly*. Even if desire subjectivism is the correct theory of welfare, S may not know that in desiring that her life goes badly, she is desiring that (some of) her desires are frustrated. But we have already ruled out such an indirect move as incoherent for judgement subjectivism.

The paradox of judgement, then, comes about directly. And this, together with the conceptual nature of judgements about one’s own good that we have already outlined, opens up a response that is not available to the desire theorist.

There are parallels between the paradox of judgement and the ‘preface paradox’ (Makinson, 1965). The preface paradox, notes that one will believe in an authored work of each claim that it is true. Yet one may in a preface acknowledge that – as is the case in any large work – at least one claim is likely to be false. One’s *general* belief about one’s beliefs as a set clashes with each individual belief. Similarly, one’s *general* judgement about the value of one’s judgements as a set clashes with each of one’s individual judgements.

Unlike in Bradley’s paradox of desire, then, there is something paradoxical *about the very judgement itself* involved in the paradox of judgement. One way of putting this is that, for anything that is the object of a welfare judgement, P believes that it is both overall fundamentally good for her, and in another way overall fundamentally bad for her. The appearance of paradox can be diffused – that is, the apparently clashing judgements can be made sense of – only if we say that G takes the form of a kind of general rule. That general rule will look something like this:

**Conditional Welfare Judgements**: It would be good for me if x happened *only if* x is part of an *even set of realised judgements*.

This is applied as a caveat to each welfare judgement (‘This would be good for me, but only if…’). This resolves the paradox. However, it does so at an odd cost to the second-order judgement itself. For it turns out, on this view, that although G is coherent, it can never be both realised *and* good for P.

In the case where P has an even number of judgements realised, G is also realised. But it is not realised as part of an even-numbered set of judgements and so, according to Conditional Welfare Judgements, is not good for P. In a case where P has an odd number of judgements realised, G has not fallen prey to the caveat; but it is also not true, and so cannot contribute to P’s welfare. As such, the paradox of judgement does not cause the same problems for judgement subjectivism as the paradox of desire did for the desire theorist.

**4. The Infant Problem strikes back**

In adding a further condition to the kind of judgments that are welfare-grounding, we add to the tension generated by the Infant Problem. It seems extremely unlikely that young infants, for instance, judge things to be not only good for them, but *intrinsically* so.

The Infant Problem, then, looms particularly large for a theory that is sufficiently complex to avoid the Paradox Problem. In general, I would suggest, subjectivists should want to be able to say that any creature that can experience pleasure and pain can have a level of welfare, on the grounds that pain is typically bad for the experiencing subject and pleasure typically good, even if by adopting subjectivism we reject hedonistic theories of welfare. But if some creatures that can experience pain cannot make judgments about their own welfare, judgment subjectivism must claim that those creatures’ pain is not bad for them, and that their pleasure is not good for them. While there is some scope to say that not all pains are bad, and not all pleasures good, this is a significant theoretical cost, especially since many of the pains that we do tend to consider uncontroversially bad for the subject are basic ‘animal’ pains of the body, and so experienced by the groups under consideration.

Dorsey considers (although does not endorse) the thought that there may be a particular kind of feeling associated with judging that something is good for us, noting Street’s (2008, 239-40) analogous analysis of the “idea of one thing's being a reason for another [which] is given by our knowledge of what it is like to have a certain unreflective experience—in particular, the experience of various things in the world as ‘counting in favor of’ or ‘calling for’ or ‘demanding’ certain responses on our part”. If this is right, there may be scope to suggest that there is also something it is like to have an experience of things as counting in favour of an action *for one’s own sake*, such that one can be said to have at least a proto-belief that an act is good or bad for you. Perhaps non-human animals do not make conceptually or cognitively complex judgments about their own good; but as De Grazia (2009: 205) puts it, “even if we don’t attribute to animals full-blown desires and beliefs, which are concept laden, we may in good epistemological conscience attribute to them *proto-desires* and *proto-beliefs*...The idea is that these mental states, though not conceptual, nevertheless have *content*.”[[6]](#endnote-6) If animals have the kinds of experiences that Street cites and, in addition, that they are able to distinguish in some proto-sense between something as ‘counting in favour of’ an act or state of affairs on the basis of their own good, and on the basis of other reasons such as the interests of their young, then there is scope to argue that animals are capable of the minimal kinds of judgments on the basis of which judgment subjectivism ascribes welfare.[[7]](#endnote-7) Such a view would also undermine Lin’s assumption (ibid. 4) that “valuing something involves having a favorable conative attitude toward it that is more psychologically complex than desire”.

Gennaro (2009, 189) suggests that there may be “a continuum of self-consciousness”, of which almost all animals are likely to have some degree,[[8]](#endnote-8) while De Grazia argues that even basic intentional action by non-human animals requires “at least some rudimentary awareness of oneself as persisting through time”, thus claiming (op cit, 209) that “an enormous range of animals are self-aware in some way and to some degree”. Relatedly, Simmons (2009, 380) notes that “concepts can vary in their degree of detail”, and considers the possibility that animals have at least a “minimum concept” of themselves as continuing subjects of experiences”, which supports the view that “animals’ displays of fear and self-protective behavior in the face of threats to their lives is evidence of their having a desire to live. Perhaps, then, a minimal form of the judgment that ɸ is good *for you* (as opposed to good *simpliciter*) requires only that you experience ɸ as in some sense ‘counting in favour of’, as Street puts it, where the reasons for that judgment relate to benefits to oneself conceived in a very minimal way, as De Grazia, Simmons, and Gennaro suggest. The most promising version of this response would show that experiencing pain typically involves some minimal conceptualization of that experience as being bad *for the experiencing agent*, such that pain is almost always *prima facie* bad for the agent. Street (2006: 144-152), for instance, suggests that although it is conceptually possible to separate the sensation of pain from a negative evaluation of that sensation, the two are tightly bound together such that creatures that feel pain almost always do have an “unreflective evaluative reaction” toward it. Street’s description of this evaluative reaction, however, only concerns the painful experience as to be avoided, which does not necessarily involve evaluating it as bad *for* the subject, let alone as intrinsically or constitutively bad for them. So while this is a possible response to the Infant Problem, it depends on a great deal that is uncertain about infant and animal cognition.

**5. Return of the subjectivist: A disjunctive view**

This response is thus a precarious one; while it is at least conceptually possible that newborn infants and animals are capable of the kinds of judgement required – albeit in a less cognitively advanced sense – this is empirically uncertain territory, and will certainly seem an unlikely claim to many. One further option is to bite a bullet and accept that some individuals to whom we normally ascribe welfare do not in fact have it. If the scope of this denial is fairly wide, this is an unattractive option, since it involves denying that the pain of a great many individuals is bad for them.

On the other hand, we could propose a distincttheory of welfare for individuals that cannot make the requisite judgments, reserving judgment subjectivism for cognitively developed creatures, and hence having a disjunctivetheory of welfare. Some versions of this move are also fairly unattractive, at least for a thoroughgoing subjectivist. For instance, we might turn to a form of hedonism for infants. Two problems arise here. First, if pain is objectively bad for infants, it is not clear why it should not also be bad for adult humans. This problem is a generic one for any hybrid view, and it is the basis of Lin’s rejection of a great many sophisticated subjectivist views, so I will address it in detail later in this section. The second problem, more specific to hedonism, is that the hedonism described here seems to be an *objectivist* view. So, at least for those subjectivists who are attracted to their view because of some of the concerns outlined in Section 2, this apparently partial concession is really to give away the store.

Instead, judgment theorists should adopt a *disjunctive* *subjectivist* view, where the welfare of infants, animals and those with some severe disabilities depends on different subjective mental attitudes than the welfare of cognitively developed adults. Lin considers the idea that we might respond to the Infant Problem by adopting such a bifurcated view, but insists that this would entail that “adult welfare diverges from neonatal welfare in a way that cannot plausibly be explained” (358). Moreover, such a view would still have infant welfare depend on some other subjective attitude, such as desire. Since Lin believes that infant cognition rules out *all* forms of subjectivism, he does not believe that such a view can solve the problem. Finally, we adopted complex subjectivism because of the Paradox Problem. Might returning to a simpler attitude – even if only partially – bring this issue back? In the remainder of this section I will outline a disjunctive view, and explain how it stands up to these worries. One version of a disjunctive view in this mould is Disjunctive Subjectivism:

**Disjunctive Subjectivism**: The welfare of individuals is determined by either (i) their judgements of whether something is intrinsically good or bad for them, or, if they are incapable of making such judgements, (ii) their desires.

I take the kind of simple desires in condition (ii) to be equivalent to Street’s description of cognitively less complex individuals making unreflective evaluations of certain things as ‘to be avoided’ or ‘to be pursued’. These, I assume, are very basic kinds of judgement; they do not involve any significant conceptual or cognitive complexity on the part of the subject, and they certainly do not involve any reference to the self, or the issue of intrinsic and instrumental value, both of which are required by the more sophisticated judgement in condition (i).

As Lin (2017, 362) suggests, it also seems plausible that at least for some cognitively undeveloped subjects, these desires are purely ‘experiential’, i.e. “a desire for an experience that is based solely on the felt quality of that experience”. If infants and other individuals are extremely cognitively undeveloped, it is unlikely that they will be capable of desiring not only, say, certain cuddling sensations, but also that they have a particular cause, such as their mother.

One worry can be dealt with succinctly. Any individual that falls on the ‘desire’ side of the disjunction will not also not be able to generate the Paradox Problem. An individual that is not cognitively advanced enough to form the judgement that something is good or bad for them is, by definition,incapable of forming the desire that things go badly for them, since that desire necessarily involves the judgement that something is bad for them. Disjunctive Subjectivism thus meets one requirement of a plausible subjectivist theory, which is to avoid the Paradox Problem for all individuals. Cognitively complex individuals, who fall into track (i) of the theory, avoid the problem because what is good for them depends on judgements of intrinsic goodness that do not fall into the paradox, as explained in Section 3. And less complex individuals avoid the problem because they are incapable of the kinds of judgements that are required for the paradox to emerge.

In the remainder of this section, I answer two challenges from Lin. Section 5.1 explains why, *pace* Lin (2017, 358), Disjunctive Subjectivism does not have adult and neonatal welfare diverge “in a way that cannot plausibly be explained”. Section 5.2 then addresses arguments against the claim that infant welfare can solely depend on even a straightforward attitude such as experiential desire. Finally, Section 5.3 considers the further issue of individuals who fall in between the two extremes of having only experiential desires, and being capable of complex value judgements.

*5.1 Adult and infant welfare*

As I have already noted, a disjunctive view of welfare must explain why two classes of individuals have their welfare depend on different factors. For instance, the view that infant welfare is hedonistic, while adult welfare is subjective, runs immediately into the concern that adults too feel pleasure and pain; if these are respectively good and bad for infants, we will need to provide some reason for claiming that they are not similarly good and bad for adults.

Disjunctive Subjectivism faces a parallel issue. The view claims that infant welfare is determined by the satisfaction or frustration of experiential desires, while adult welfare is determined by a rather more complex attitude, judgement that something is intrinsically good or bad for the subject. What is more, Disjunctive Subjectivism *denies* that the satisfaction (frustration) of experiential desires is necessarily good (bad) for adult subjects. So we are left with a parallel to the claim that troubled disjunctivists about hedonism: why should the satisfaction of experiential desires be good for infants, but not for adults?

Lin (2017, 359) considers one potential explanation: when a human being develops the relevant cognitive capacities, judgement subjectivism *becomes* true of her because she is now “an importantly different kind of creature: one capable of having a *conception of her own good*”. My view is that this is indeed the obvious, and correct, explanation. But Lin suggests that while developing a new capacity might explain an *increase* in the kinds of things that can determine your welfare, it is not plausible that it could explain a *decrease* in the way that Disjunctive Subjectivism supposes. After all, it is not as though adults, in developing the capacity for critical judgements about their own welfare, also lose the capacity to have experiential desires.

It is important to be clear what this objection amounts to. For instance one might, mistakenly, express this idea in the following principle:

**Welfare Capacity**: Developing a new capacity cannot restrict the range of things that can determine your welfare.

Welfare Capacity is obviously false. Consider, for instance, Agar’s (2013) worries about the development of radical biomedical enhancements that could increase our cognitive capacities vastly beyond the current range. Agar suggests that one problem with this idea is that it will alienate us from things that we currently value. The work of Austen or Beethoven, for instance, will now be far too simplistic for us to enjoy; we will crave more intellectually stimulating work. One might dispute the details of Agar’s argument. But I do not think that it is obviously incoherent in the way that it would have to be if Welfare Capacity were true.

However, Lin’s position is not captured by Welfare Capacity. His claim is not that developing a new capacity cannot undermine something as a source of value; rather, he insists that:

**Welfare Capacity\*:** Developing a new capacity cannot restrict the range of things that are *basically* good for you.

Something is basically good for you, says Lin (2017, 354), if it is good for you “not solely in virtue of being suitably related to other things that are good for you”. And, he might insist, reading Tolstoy or Atwood is good for you only because it brings you some other good, such as (a distinctive kind of) pleasure.

Welfare Capacity\* is not obviously false in the way that Welfare Capacity is. But it is still false. The reason it is false is that the development of a new capacity can have implications for the ‘output’ of a former capacity, because the new capacity involves a kind of higher order critical reflection. Consider, for instance, Korsgaard’s (1996, 90-97) suggestion about the difference between rational and non-rational beings. Rational beings are capable of moral obligation in part because they are capable of critical reflection; rather than simply pursuing inclinations, we can also step back and critically assess our inclinations, asking ourselves whether they *really* provide reasons for us to act.

The development of a capacity for critical reflection can, then, reduce the range of things that it is *morally permissible* to do. When an infant lashes out in anger, this action is not morally impermissible. When an adult lashes out in anger, this may be morally impermissible because the adult has, and the infant lacks, the capacity to reflect on their desires and whims, and to decide not to act on them.

At a basic level, then, we might turn Lin’s argument back on him. Since the development of a cognitive capacity can plausibly make a difference towards what we are morally permitted to do, why *shouldn’t* it also make a difference to what is good for us? But while I think that this is at least a reasonable question, it is hardly a decisive one, and leaves us at something of an impasse.

More forcefully, though, I think that we can offer an equivalent explanation for why adult welfare is determined by something different than infant welfare is. Just as our ability to critically reflect on our possible actions makes a difference to the moral status of those actions, so too does it make a difference to their welfare implications. Infants cannot take a critical stance on their experiential desires. What is more, to desire something is to extend a kind of basic endorsement towards that thing. When infants have experiential desires, then, they are extending *the most complex kind of endorsement of which they are capable* towards the object of those desires. When an adult has an experiential desire, on the other hand, they are capable of more complex forms of critical endorsement or disapproval. As such, the mere fact that an adult has an experiential desire is not enough to dictate whether the object of that desire is good for them, because we also need to know what their assessment of that desire is.

Benjamin Yelle (2016) offers response to Lin that is similar to this, but with several important differences. First, Yelle suggests that infant welfare is hedonistic. As I have suggested, if we understand hedonism as an objectivist view, this is not a response that a thoroughgoing subjectivist should be comfortable with.

Second, Yelle’s argument is based (following Rosati, 1996) on the view that “the good of a creature…must suit its nature” (1411). And while our nature as newborns is an *animal* nature, our nature for most of our lives is as *persons*. This latter stage, says Yelle, centrally includes the capacity for rational self-governance. This aspect of the view faces the same problem as the first, though more complexly. Yelle argues that we have both pragmatic and theoretical reasons to focus on autonomy in our characterisation of our welfare *qua* persons. The pragmatic reason is that it allows us to accord people the respect they are due as valuers; but this is not related in any obvious way to the question of whether a theory that does this is a correct *description* of welfare. The theoretical reason is that it allows us to “accommodate more of our intuitions about welfare”. But Yelle then goes on to mention a further motivation, i.e. the appeal to Rosati’s view.

The appeal to our ‘nature’, though, has considerable problems. First, it should be noted that the appeal to our intrinsic nature is characteristic of *perfectionist* views of welfare. In particular, if our welfare is to be determined according to our membership we are to be categorised according to our membership in a ‘kind’, this at least potentially opens up the possibility of there being *other* features that are also distinctive of our kind, and which compete with autonomy as the overriding feature in a way that moves us away from subjectivism.

Second, although Yelle uses the appeal to kinds to endorse a view of the valuer as in some sense sovereign, it is surely at least a possibility that someone might *fail to value their capacity to value*.[[9]](#endnote-9) In such a situation, Yelle’s view either seems to suggest that we should not take their evaluative judgements seriously (since respecting their autonomy *in this instance* involves agreeing with them that their capacity to value is not distinctively special in the way that other persons’ capacities are), or insisting that they are *wrong* in this evaluative belief because it is constitutive of the kind of creature they are that they have reason to value that capacity. The former response suggests – in line with my argument – that what matters is fundamentally the individual, rather than any kind she might belong to; the latter response is not a subjectivist one.

Finally, Yelle’s argument implies that if it were *not* constitutive of the kinds of beings we are – that is, if complex value judgements played a more peripheral role in our lives – then hedonism might well be true of us as well. I am not sure exactly what makes something definitive of a kind; but there are presumably activities that we engage in occasionally – perhaps even often – that are not so constitutive. If value judgements played a more peripheral role in our lives, it would still be constitutive of *those judgements*, and the capacity to make them, to be a ‘step back’ from our instinctive desires and urges, and to assess them for their potential as reasons.

My argument is therefore different to Yelle’s in being more strictly subjectivist. Adults’ welfare is to be assessed by reference to their judgements not because of any kind-membership they possess, but simply because – in the absence of objective value that the radical subjectivist assumes – it is the nature of *value judgements themselves* that determines their role in grounding welfare*.* Lin might, however, suggest that this response is incomplete. For while the ability to critically endorse or disapprove of something might imply that experiential desire *sometimes* doesn’t determine desire, it does not explain why it should *never* determine desire, as seems to be demanded by a genuinely disjunctive view. In particular, Lin suggests that we cannot explain why experiential desires do not continue to determine the welfare of an adult who “acquires the capacity to have welfare beliefs at the normal stage in her development, but who fails, for a period of time after acquiring the capacity, to actually believe anything to be basically good for her” (2017, 359). Such an individual is capable of more complex critical reflection on her desires, but does not actually engage in such reflection. Lin suggests that such a view would imply that a subject can go from extremely high welfare to a welfare level that is at best neutral, without engaging in any change of attitude about any object. And this, he thinks, is implausible.

One thing to say about this objection is that it seems to assume that you only count as making a welfare judgement if that judgement is explicitly formulated in terms of a theory of welfare. Lin cites, for instance, the fact that G.E. Moore (1903, 98-99) claimed not to have any welfare beliefs because he could not understand how something could be good for someone, rather than good *simpliciter*. But it is not clear that a person who – perhaps for theoretical reasons – does not believe in welfare judgements at a philosophical level (or even who simply does not think about things in philosophical terms) cannot nonetheless count as *making* such judgements.

Imagine, for instance, a person who is offered a cigarette, but refuses on the grounds that ‘cigarettes cause cancer’. It hardly seems a required that this person has a worked out theory of welfare in mind for us to judge that their refusal is grounded in the view that cigarettes are bad for them. So the individual Lin describes – that is, the person who could but does not make welfare-related judgements – is not someone like Moore, but someone who only ever acts on the basis of entirely unreflective desires.

In addition, it is still true of such an individual that their experiential desires do not constitute the most complex form of endorsement of which they are capable. So an individual who forms no complex evaluative judgement about an object of desire, but is capable of doing so, is doing something that an individual who simply *cannot* form such judgements is not: they are *failing to endorse* that object in as complex a way as they might. Again, there are parallels with Korsgaard’s account of obligation. Even if I never do reflect on my actions, the fact that I am capable of doing so makes me morally culpable in a way that a creature that cannot so reflect is not.

This may seem to have a further troubling implication, however. Most people do not engage very often in in explicit judgements about the contribution of various goods to their welfare. So this view may seem to imply that most people’s welfare is neutral most of the time. Similarly, there are some situations that are surely bad for us, but which are such that we are unlikely to form explicit welfare-based judgements about them. If one is being tortured, for instance, one will certainly want the experiences to stop. But one is unlikely to further think that it would be good if the torture stops *because it is having a negative impact on one’s welfare*.

In response to this, I will make one clarification, and one further adjustment to the theory. The clarification is to insist that the latter objection overestimates the complexity of the attitudes involved. While it is true that one is unlikely to form an independent belief that torture is a negative contributor to welfare while it is happening, one’s experiential desire may well involve a more complex evaluative stance than the equivalent experiential desire of a young infant. When an adult is undergoing severe pain, for instance, they may not only want the pain to stop; they may also regard it as ‘horrible’, ‘awful’, ‘excruciating’. These evaluative terms are not explicitly about welfare; but when applied to one’s own experiences, they imply some kind of consideration of one’s own welfare. There is a difference, for instance, between describing someone else’s suffering as excruciating, and describing my own suffering in those terms. When I so describe my own experience, I express an internal view of things that relates to a sense of my own welfare. And while that is a difference I may not be able to articulate under torture, my evaluative capacities mean that it is contained within my judgement.

My concession, however, is to acknowledge that the most plausible form of Disjunctive Subjectivism may rely not only on actual welfare judgements, but may also rely on the implications of those judgements. In other words, the most plausible form of the theory will involve an appeal to something like Sharon Street’s ‘Humean constructivism’. In Street’s (2012) terms, Humean constructivism says that what is valuable depends ultimately on a valuing individual’s evaluative stances, but need not depend *directly* on those stances. For instance, if someone holds a negative view of torturing humans solely because they think that pain is bad for people, but does not hold a similar view of torturing animals, we can nonetheless say that they are *committed to* the view that torturing animals is bad because it is implied by their more basic attitude toward torture and pain.

Similarly, even if one does not make an explicit *welfare* judgement about a particular object or event, some implication for one’s welfare may be implied by other attitudes one has. This explains how a Disjunctive Subjectivist can say of a particular person that it *would be* good for them to φ, even where φ-ing is not something they have ever considered.

Someone who fails to make an explicit welfare judgement, then, may still differ from an individual who cannot make such judgements in a number of salient ways. First, they may have made other judgements that have potential implications for the case at hand. Second, the fact that someone fails to endorse or disapprove of something where they are able may have an evaluative relevance that a simple inability to endorse or disapprove does not. Third, someone who has a capacity to form welfare beliefs may form them without explicitly adopting a theory of welfare, and without consciously employing the concept of welfare (just as we can form beliefs about chairs without consciously employing the concept ‘chair’).

The conclusion of this discussion, however, is that Disjunctive Subjectivism should be revised and clarified. It now stands as such:

**Disjunctive Subjectivism\*:**  The welfare of individuals who are capable of such judgments is determined by (i) their direct judgements (including implicit judgements) of whether something is intrinsically good or bad for them and the evaluative implications of those judgements or, if they are incapable of making such judgements, (ii) their desires.

*5.2 Infant desires*

I have explained why the welfare of cognitively complex individuals can plausibly be determined by very different considerations than the welfare of cognitively simpler individuals. The second objection to my theory that comes from Lin’s argument is the idea that it is implausible that neonatal welfare is determined by experiential desire alone.

Lin’s argument that desire subjectivism cannot be true of infants rests on the following suggestion. First, as I outlined earlier in the section, it seems at least plausible that at least some infants can have only experiential desires. An experiential desire related to a neonate’s mother cuddling them can only be for the particular feeling associated with the cuddle; it cannot be for their mother in particular to cuddle them. While being cuddled by a person who is not their mother might be experientially different, so that not just any cuddle will do, this does mean that their desire will be met just as well by an excellent facsimile of their mother’s cuddle, such as one might have in a sophisticated experience machine.

Lin then suggests (2017, 362-363) that a child whose neonatal period is spent with actual parental contact (Adam) has a higher welfare level during that time than a child whose neonatal period is spent in an experience machine, reproducing that contact exactly (Bill). Yet desire subjectivism suggests that their welfare is the same during this period. It is important to note that Lin’s argument here is restricted solely to questions about relative levels of welfare *during the neonatal period*. He allows that desire subjectivism will correctly judge that the two individuals’ entire lives differ in quality, because at some point they will develop desires that are not purely experiential. Once Bill starts to desire not simply certain experiences, but actual contact with his parents, or to have friends, desire subjectivists can say that his desires are not being met. But Lin insists that before the children are capable of non-experiential desires, their welfare is still different. He calls this claim, ‘Neonatal Period’:

**Neonatal Period:** Adam’s total well-being during the neonatal period is at least somewhat higher than Bill’s.

The plausibility of Lin’s argument against desire subjectivism for infants, then, rests squarely on the claim that a newborn infant in the real world has a better life than his experience machine twin. This argument depends in turn on several claims.

First, Lin notes that if you were entrusted with the task of looking after a child who will soon be born and who will not survive beyond the neonatal period will have the best possible life, you would not be indifferent between actually caring for them, and placing them in a reliable experience machine, even ignoring any considerations other than those relevant to welfare. If that is accurate, he says, “then either you believe Neonatal Period or you believe that it *might* be true” (2017, 363).

The stipulation that we ignore considerations other than those relevant to welfare is quickly made, but important. For there are many reasons that we might be uncomfortable about the idea of placing a baby in an experience machine rather than caring for it. We may feel, for instance, that a duty of care involves certain actions quite apart from their effect on an individual’s welfare. We might think that placing a child in the care of a machine shows a level of disrespect to that child that is unjustified. Finally, we may simply feel squeamish about the idea of leaving a child in the care of a machine, even if we have stipulated that it will be safe and happy. Lin asks us to ignore features of the case that, even if we are happy to admit that they are not theoretically relevant, nonetheless unavoidably *strike us at an emotional level*. In short, my response to Lin is that relying on intuitive reactions to such an artificial example is a rather weak argument for the claim that infant welfare is determined by desire.

Lin has (2016, 321) responded to this kind of suggestion by insisting that we are perfectly capable of distinguishing between influences on our overall judgement from considerations about welfare, and those from considerations about non-welfare related factors. He adds that, of someone who has such an intuition, having tried to focus only on welfare-related values, “It would be unreasonable to doubt that their intuition is about welfare just as it would be unreasonable to have analogous doubts about those who claim to have the intuition that if two lives are exactly the same except that the second contains more pleasure, the second life is higher in welfare”.

There are a few things to say in response to this. First, this equivalency might be reasonable in cases of comparing adult welfare. But even hardened philosophers are capable of being swayed by additional concerns when it comes to the welfare of children. To insist that all such intuitions are of equivalent ‘neatness’, even though we can point to psychological factors that might confound one but not the other, is itself unreasonable. For instance, there may be evolutionary reasons for us to be more uneasy about the idea of ‘abandoning’ a child to the care of a machine. I do not disagree with Lin that people may be able to set such considerations aside; what seems to me rather less warranted is to both acknowledge that such biases are possible, and to take even a *slight* inclination in favour of looking after a child in person as evidence that you think Neonatal Period is true.

Admittedly, Lin does somewhat caveat the strength of his conclusion, by suggesting that if we are not indifferent between the two options we either believe that Neonatal Period is true, “or you believe it *might* be true”. But I am not sure what ‘might’ is supposed to convey in this passage. After all, although I am arguing for welfare subjectivism, I think that objectivism *might* be true, in the sense that I acknowledge it as a possibility. Yet this alone does not offer much reason to believe in welfare objectivism. On the other hand, ‘might’ may take on a stronger meaning here than ‘merely possible’. Perhaps Lin intends ‘might’ to mean something like ‘has a strong chance’. That, however, seems to me an unwarranted conclusion for the reasons outlined above.

Second, it is important, albeit fairly obvious, to note that what we may have here is a simple clash of intuitions. Even if we grant that someone – particularly someone who is used to abstracting away from different factors, such as Lin himself – might be able to focus only on welfare-related factors of the case, and still come to the conclusion that Adam is better off than Bill, we might equally note that some people might not have such a response. Lin suggests that we must be careful not to allow our theoretical commitments to sway our ‘intuitive’ response; but it should presumably be no more reasonable to reject such a response on the grounds that it ‘probably’ derives from a theoretical commitment than it is to reject a response because it may have been influenced by non-welfare factors.

Finally, as Lin himself notes, our intuitive reaction to the case of Adam and Bill is hardly a conclusive argument against desire subjectivism. If there were a theory of welfare that had no counter-intuitive implications, we might well all have agreed on it by now. So even if we acknowledge that the implication of desire subjectivism for the case of Adam and Bill is counter-intuitive, this may only give us a weak reason to reject desire subjectivism.

Lin’s second argument for the truth of Neonatal Period is the claim that Bill is properly an object of pity in a way that Adam is not. But, he says, “to pity someone is to feel bad for him on account of the fact that he is, in certain respects, badly off” (363). So if our pity for Bill is warranted, he must be badly off.

Let’s allow that if we were to confront a case like Bill’s, we would feel an emotional response; I might well feel a particular negative emotion, directed at wanting to remove Bill from the experience machine and care for him, or a wish that he had not been placed there. There are two further questions to confront. The first is, is that response pity? The second is, is that response warranted?

The question of whether the emotional response that we feel is pity depends on whether it is essential to pity that we be moved by a judgement that someone is badly off in some way. Lin insists that this is simply what pity is. Assume, on the one hand, that he is incorrect, and pity should be defined solely by a subjective phenomenological feeling. Then, my reaction toward Bill’s case might be pity, but might not back up Lin’s argument at all. On the other hand, suppose that we accept Lin’s definition of pity; pity essentially involves a judgement that someone is badly off. But now Lin cannot insist that, if I feel the emotion described above, *that must be pity*. For we have defined pity as being determined not by the particular ‘feel’ of an emotion, but by the judgements that cause it. And since – as Lin seems to acknowledge when he says in his first argument that we must screen out non-welfare considerations – there may be *other* reasons for such a feeling, we cannot know merely from experiential data whether what we are feeling is pity.

So, when Lin says that Neonatal Period is backed up by “the observation that…Bill is *pitiful* in a way that Adam is not”, what this amounts to is saying that Neonatal Period is backed up by the judgement that Bill is worse off than Adam. For only if we *already* think Bill is worse off than Adam can any emotion reaction we might have toward the case be described as pity. But this is question-begging, since Neonatal Period is simply the claim that Bill is worse off than Adam. Saying that Bill is intuitively pitiful, in a way that Adam is not, is essentially to say that Bill is intuitively worse off than Adam in some respect.

The second question is whether, even if it is pity that I feel, that pity is warranted. Lin suggests that “if the pity that we feel toward Bill is *warranted*, then he really is in certain respects badly off”. This is true if pity necessarily involves a judgement that Bill is badly off. But this is a rather misleading truth, since while it appears to be giving additional support to Neonatal Period, it does no such thing. All it does, in fact, is restate the definition of pity with which Lin is working. It might be true that *if* our judgement of pity is warranted, then Bill is worse off than Adam. But the question of *whether* that judgement is warranted depends precisely on the truth of Neonatal Period. So we can hardly use the suggestion that our pity judgement is warranted to back up that claim. The appeal to pity, then, adds nothing to Lin’s argument as it begs the question in favour of Neonatal period.

I suggest, then, that the second branch of disjunctive subjectivism remains plausible in the face of Lin’s criticisms. Since both branches – and the idea that welfare might be determined by different factors for cognitively different individuals – the view is a plausible theory of welfare.

*5.3 Intermediate cognitive complexity*

The final issue with disjunctive subjectivism that needs considering is the case of what I will call ‘intermediate cognitive complexity’. Disjunctive subjectivism says that unless an individual can make certain complex judgements, their welfare depends on their desires. I have mainly contrasted such complex individuals with those at the opposite end of the spectrum, i.e. individuals who are capable only of experiential desires. But it is important to note that humans do not move straight from being capable only of experiential desires to being capable of complex judgements about their own welfare. There seems to be at least one intermediate stage, namely the stage where we can desire *specific* *objects*, rather than just experiences, but cannot yet have a conception of our own welfare, or differentiate between intrinsic and instrumental contributors to welfare even in a fairly simplistic way.

Consider, for instance, an infant who has passed the neonate stage so that they can form desires for particular objects in front of them, but is not yet developed enough to make judgements about welfare. Disjunctive subjectivism says that their welfare is determined by their desires.

But since this infant is now capable of forming desires for objects, rather than just experiences, they are clearly capable of forming desires for objects that seem to be clearly not good for them.[[10]](#endnote-10) For instance, a baby might see a shiny object, and form a desire to place it in its mouth. If that shiny object is a small bead that will cause the child to choke to death, satisfying that desire seems not to be good for them.

There are two possible responses to this. The first is to complicate the desire-based disjunct of disjunctive subjectivism, so that only desires of a certain kind (e.g. what one would want if they were ideally rational, or knew all the consequences) can affect welfare. I think this is a rather difficult adjustment to make in the case of infants, since there are clearly some things that are good for babies that a fully rational being would at least be unlikely to desire (e.g. playing with colourful wooden blocks). In addition, such idealizations seem, as Heathwood (2014, 40) puts it, to “abandon the core ideal of subjectivism”, which is that what is good for you must be connected to “*you yourself*” (see also Murphy, op cit).

A second response might make use of Lin’s observation that the acquisition of more complex cognitive traits does not remove the existence of experiential desires, specifically the desires for pleasure, and for the avoidance of pain. In many cases, a disjunctive subjectivist can insist that getting the shiny object into their mouth *is* good for them, but note that the result (which, remember, was not desired) is bad for them because it *prevents* the satisfaction of certain other desires. For instance, the experience of choking will be painful and distressing, and so clearly bad for the child. However, this response is also incomplete. For we can change the example so that whatever it is that he child forms a desire for will kill them instantly but painlessly.

But the disjunctive subjectivist has available a move analogous to the adjustment we made, in Section 5.1, to the first part of the theory. This adaptation would be to say that it is good for infants to get what they want, *and* for them to have experiences that they *would* want to continue (and, respectively, bad for them to have experiences that they *would* want to end) if they were to have them.[[11]](#endnote-11) Notice that this is related but importantly different to Heathwood’s (op cit.) solution to the related problem of *uninformed* desires (see also Hubin, 1996). Heathwood suggests that, if my desire for cherry pie will kill me because I am unwittingly allergic to cherries, we can appeal to my “very strong desire not to be sent to the emergency room”. But an appeal to what someone would come to desire if they were to experience it still retains the important connection between welfare and the *actual* person, rather than some idealised version of them.

This extension is well motivated. After all, the desire theory claims that prudential value depends on certain facts about the subject – her values and preferences – that determine what she desires, and explain why getting what she wants is good for her. But these facts also have other implications, including implications for what the subject would evaluate positively or negatively if she knew about it.

One might argue that even if continued life *would be good* for the child, painless death is not bad for them because although they do not get what they would desire, they also do not experience anything that they do not, or would not, desire. This does not, of course, mean that we can be indifferent to whether young children suffer painless deaths. The fact that continued life would be good provides us with a strong welfare-based reason to keep them alive; what is more, it is also possible to insist that people have a right to life even if death would not be bad for them.[[12]](#endnote-12) But this is simply an instance of broader debates about whether death is harmful for us (e.g. Bradley, 2009), and so not a distinctive problem for subjectivists.

This amendment, then, allows us to acknowledge that there are some things that would be good for young infants that they are not capable of desiring because they are too fundamentally different from experiences they have had until now.[[13]](#endnote-13) It also gives us a possible response to the case of painless death. Although painless death is not directly bad for an infant, it may be *indirectly* bad since it prevents them from having and satisfying new desires that would have been good for them.

This is in some ways similar to Simmons’ (2009) response to a related question. Simmons notes first that even for developed adults, many of our desires are *dispositional*. That is, it seems inaccurate to say that I want to live only when I am consciously entertaining that desire; someone who kills me in my sleep does something that I do not want them to do. Simmons argues that there is a constitutive relationship between enjoying something, and having a desire for it. As he puts it, “One has a feeling of care toward the thing one enjoys, in such a way that one is disposed or motivated in one’s behaviour to pursue that thing. It is in this sense that a being’s enjoyments count as desires.” This is because enjoyment should not be seen as a “temporary experience”; in many cases to say that someone has an enjoyment implies “that one has a continuing tendency to experience enjoyment” over that thing. As such, Simmons argues, we can appeal to animals’ and infants’ *present* (dispositional) desires to explain why even painless death is bad for them. Although he does not explicitly say so, it seems clear that this view is compatible with a subjectivist approach to value.

*6. Conclusion*

I have argued that disjunctive subjectivism is a plausible view even in the face of Lin’s criticisms. Disjunctive subjectivism says that a person’s welfare is affected by their direct judgements about what is intrinsically good for them or, if they are incapable of such judgements, getting what they want, or getting things that they *would* want if they had experienced them before.

I suggested that this shift in welfare-grounding attitudes is plausible because the development of more complex capacities *overrides* the welfare-grounding nature of desires. I also considered some potential objections from Lin to the desire-based disjunct. Even though we would clearly be reluctant to place a neonate in a sophisticated experience machine rather than look after it ourselves, this does not show that such a child would actually have a worse life.

Disjunctive subjectivism thus avoids the Infant Problem, because it does not claim that infants (or other individuals who are not capable of cognitive complexity) do not have welfare levels. It also avoids the Paradox Problem because individuals are either not capable of forming the problematic desire that their life goes badly, or because they lie on the side of the disjunction that makes such desires irrelevant. As such, those who think that welfare is determined by our subjective mental attitudes have three (broad) options. They can deny that infants, animals, and adults with severe mental disabilities are capable of having welfare. They can reject one of the assumptions supporting the paradox of desire

**References**

Agar, Nicholas. (2013) *Truly Human Enhancement: A Philosophical Defence of Limits*. Cambridge: MIT Press

Bradley, Ben. (2007) ‘A paradox for some theories of welfare’ *Philosophical Studies* 133: 45-53.

(2009) *Well Being and Death.* Oxford: Clarendon Press.

Brandt, Richard. (1972) ‘Rationality, egoism, and morality’, *The Journal of Philosophy* 69: 681–97.

Bruckner, Donald. (2010) ‘Subjective well-being and desire satisfaction’, *Philosophical Papers* 39(1): 1-28

DeGrazia, David. (2009) ‘Self-awareness in animals’, *The Philosophy of Animal Minds,* edited by Robert Lurz, Cambridge: Cambridge University Press. pp.201-217

Dorsey, Dale. (2012) ‘Subjectivism without desire’, *Philosophical Review* 121: 407-442.

Feldman, Fred. (2004) *Pleasure and the Good Life*.Oxford: Oxford University Press.

Finnis, John. (1980) *Natural Law and Natural Rights*. Oxford: Clarendon Press.

Foot, Philippa. (2001) *Natural Goodness*. Oxford: Oxford University Press

Garner, Richard T. (1990) ‘On the genuine queerness of moral properties and facts’, *Australasian Journal of Philosophy* 68(2): 137-146

Genarro, Rocco. (2009) ‘Animals, consciousness, and I-thoughts’, *The Philosophy of Animal Minds*,edited by Robert Lurz, Cambridge: Cambridge University Press. pp.184-200

Hall, Richard J. (1989) ‘Are pains necessarily unpleasant?’, *Philosophy and Phenomenological Research* 49 (4): 643-659

Heathwood, Christopher. (2005) ‘The problem of defective desires’, *Australasian Journal of Philosophy* 83: 487-504.

(2014) ‘Faring well and getting what you want’, *The Ethical Life: Fundamental Readings in Ethics and Moral Problems*, edited by Russ Shafer-Landau. Oxford: Oxford University Press, pp.31-42

Horgan, Terence and Mark Timmons. (1992) ‘Troubles on moral twin earth: Moral queerness revived’. *Synthese* 92(2): 221-260

Hubin, Donald. (1996) ‘Hypothetical Motivation’, *Noûs* 30(1): 31-54

Joyce, Richard. (2001) *The Myth of Morality*, Cambridge: Cambridge University Press

(2006) *The Evolution of Morality*, London: MIT Press

Korsgaard, Christine. (1996) *The Sources of Normativity*.Cambridge: Cambridge University Press

Lin, Eden. (2016) ‘How to use the experience machine’, *Utilitas*. 28 (3): 314-332.

(2017) ‘Against welfare subjectivism’, *Noûs* 51(2): 354-377

Mackie, John. (1977) *Ethics: Inventing Right and Wrong*. New York: Penguin

Makinson, David. (1965) ‘Paradox of the Preface’. *Analysis* 25: 205-207

Mill, John Stuart. (1863) ‘Utilitarianism’, *Utilitarianism and On Liberty*, edited by Mary Warnock (2003). Oxford: Blackwell Publishing, pp.181-235.

Murphy, Marc. (1999) ‘The simple desire‐fulfillment theory’, *Noûs* 33: 247–72.

Nichols, Shaun (2004) *Sentimental Rules: On the Natural Foundations of Moral Judgement*. Oxford: Oxford University Press

Olson, Jonas. (2014) *Moral* *Error Theory: History, Critique, Defense*. Oxford: Oxford University Press

Parfit, Derek. (1984) *Reasons and Persons*. Oxford: Oxford University Press

(2011) *On What Matters* Volume 1. Oxford: Oxford University Press

Railton, Peter. (1986) ‘Moral realism’, *Philosophical Review* 95: 163-207.

(2004) ‘Facts and values’, *Facts, Values and Norms*. Cambridge: Cambridge University Press

Rosati, Connie. (1996) ‘Internalism and the good for a person’, *Ethics* 106(2): 297-326

Simmons, Aaron. (2009) ‘Do animals have an interest in continued life?’, *Environmental Ethics* 3(4): 375-392

Skelton, Anthony. (2015) ‘Children’s Well-Being: A Philosophical Analysis’, *The Routledge Handbook of Philosophy of Well-being*, edited by Guy Fletcher. London: Routledge, pp.366-377

Skorupski, John. (1999) *John Stuart Mill*. New York: Routledge

Skow, Bradford. (2009) ‘Preferentism and the Paradox of Desire’, *Journal of Ethics and Social Philosophy* 3(3): 1-16

Sobel, David. (2009) ‘Subjectivism and Idealization’, *Ethics* 104: 784-810

(2011) ‘Parfit’s Case Against Subjectivism’, *Oxford Studies in Metaethics* 6, edited by Russ Shafer-Landau. Oxford: Oxford University Press, pp.52-78

Street, Sharon. (2006) ‘A Darwinian Dilemma for Realist Theories of Value’, *Philosophical Studies* 127: 109-166.

(2008) ‘Constructivism about Reasons’, *Oxford Studies in Metaethics* 3, edited by Russ Shafer-Landau. Oxford: Oxford University Press, pp.207-45.

(2012) ‘Coming to terms with contingency: Humean constructivism about practical reason’, *Constructivism in Practical Philosophy*,edited by James Lenman and Yonatan Shemmer. Oxford: Oxford University Press, pp.41-59

Yelle, Benjamin. (2016) ‘In Defense of Sophisticated Theories of Welfare’, *Philosophia* 44(4): 1409-1418

1. I follow Street (2006, 135-141) in excluding certain views that are often described as ‘realist’, but which have value fully depend on our attitudes, e.g. Railton (1986). [↑](#endnote-ref-1)
2. This is not a claim about all views that can be described as ‘hedonistic’ (e.g. Feldman’s (2004) ‘Attitudinal Hedonism’, according to which pleasure is just a kind of pro-attitude), but about Mill’s brand of hedonism. [↑](#endnote-ref-2)
3. This is not to say that there is no possible form of subjectivism that rejects these assumptions in a way that would avoid the paradox. For instance, the third assumption (about the relationship between time-relative and extended welfare) could be replaced by the claim that only desires about one’s lifetime contribute to lifetime welfare. The second assumption could be rejected in favour of the claim that we can have multiple levels of welfare at any time. However, simply adding more complexity to our theory will not help. Take a theory which says that lifetime welfare is determined by time-relative welfare *and* the ‘shape’ of a life (e.g. whether it started badly and improved, or started well and declined). So long as a life’s shape contributes to overall welfare in a way that is commensurable with time-relative welfare, we would only need to adjust the numbers to generate the paradox again. [↑](#endnote-ref-3)
4. This also includes the assumption that (at least within a single individual’s set of desires) we can not only rank their desires ordinally – saying of any two outcomes which is better, or if they are equivalent – but also how much better one outcome is than another. [↑](#endnote-ref-4)
5. Dorsey (2012, 416) adds an additional stipulation to JS, which is that the relevant welfare-affecting judgments must be what the subject would judge good for them if their beliefs were “rendered coherent” collectively. However, it is not clear that one need make this stipulation to avoid the Paradox Problem. The belief that ‘my life going badly for me is good for me’ seems to be *internally* incoherent; so, all that judgment subjectivism needs to avoid the paradox is that judgments must be internally coherent, not that they must be such that they would remain in a fully coherent set of judgments. [↑](#endnote-ref-5)
6. De Grazia does want to attribute ‘full’ and not merely proto-beliefs to animals. But it might be that animals have full beliefs about some subjects, and merely proto-beliefs about others. [↑](#endnote-ref-6)
7. Even this may be unnecessary. It would be sufficient if an animal only *ever* experienced things as called for because of its own interests; on this view, an animal need not have any sense of things benefitting others. [↑](#endnote-ref-7)
8. In fact, Gennaro suggests that “almost all” animals are capable of quite complex forms of self-consciousness. [↑](#endnote-ref-8)
9. Though see Korsgaard (1992), and Street’s (2008) response. [↑](#endnote-ref-9)
10. See also Parfit (1984, Appendix I); Bruckner (2010) [↑](#endnote-ref-10)
11. Parfit (2011,82) claims that liking or disliking something is importantly different from desiring it in a way that causes problems for subjectivists. See Sobel (2011) for a response. [↑](#endnote-ref-11)
12. Indeed, people may have a right to life even if life will not be good for them. The fact that someone is suffering, and has no hope of remittance, does not give us a right to kill them without their consent. [↑](#endnote-ref-12)
13. See e.g. Sobel (2009, 336); Simmons (2009, 385). Sobel raises this as a problem with respect to adults for desire subjectivism (“…suppose that I would love the taste of pineapple if I were to taste it but now have no desire to do so. My current lack of desire for pineapple does not entail that I would not be benefitted by eating it”), but it is only a problem with respect to infants for disjunctive subjectivism. Sobel suggests that this kind of case motivates *idealized* subjectivism; but it is not clear why we should need to idealise our agent to say that she would want something if she experienced it. [↑](#endnote-ref-13)