Philosophy of Happiness Essay

Adam Chalmers

November 6, 2016

When you or I make a decision involving another person, we frequently attempt to do right by the other party. We often factor the other person's welfare into this decision — will our actions make this person better or worse-off? We may need a theory of welfare in order to make decisions about other people — a theory which explains what a person's best interests are, what makes their life good for them to live.

Scholars have proposed many theories of welfare, but frequently these theories only analyse welfare *for humans*. However, as the dominant species on the planet, humans frequently make decisions which affect non-humans. We often try to act ethically towards animals, or towards the natural environment (for example, when we consider agricultural or forestry policy). This suggests we, as human decision makers, need a theory of welfare which extends to non-humans. How can we make ethical decisions about animals or environments without considering their welfare?

Furthermore, as science and technology progress, we can expect to encounter previously unknown beings, whether artificial (such as robots or artificial intelligences) or natural (such as aliens or other spacefaring civilisations). Ideally, humans should have a theory of how to consider these creatures interests *before* we encounter them. If we leave the question of alien or robotic welfare until we encounter these beings, we may discover too late that we have already caused irreparable harm.

How are we to formulate a general theory of non-human welfare? In this paper I propose one solution to this problem by extending Daniel Haybron's work. The theory of welfare outlined in Haybron (2008) is eudaimonic and puts self-fulfilment as constituting welfare. His work focuses on happiness, which is critical to fulfilling one's emotional nature, i.e. fulfilling a very important part of one's self. However, he acknowledges other components of one's self (such as

one's deliberative-rational self and one's physical self) which are also important to welfare. I propose that welfare for beings without affective or emotional states (such as plants or robots) consists of fulfilling these components of the self which they *do* possess. This general theory of welfare would help us determine the welfare of non-humans and may thus be crucial in acting ethically towards non-human beings.

In the remainder of this section, I will introduce the philosophical notion of welfare and the various ways we use it. In section 2 I will outline Haybron's theory of welfare and identify ways it could be extended. Section 3 briefly identify some types of non-human beings and discusses how to modify Haybron's welfare theory to accommodate them.

1 What is welfare?

Roughly speaking, a person's welfare is whatever is good for them, in a prudential sense. Welfare is a measure of how good someone's life is, for the person living it. Note that "good" is used not in ethical terms (as in "the good" or "morally good"), but in prudential terms. A sadist might live a good life (in the prudential sense) by beating small animals, even though this life would not be good in the *moral* sense. Welfare is roughly similar to *utility* in ethics.

This explains the meaning of the philosophical term "welfare", but not what welfare actually consists of. Many competing accounts of welfare claim to have identified the exact nature of welfare, but there exists no philosophical consensus over which, if any, is correct.

Objective list theories claim one's welfare is a result of their acquiring several intrinsically welfare-increasing goods. For example, Gert (1998) claims welfare consists of consciousness, ability, freedom and pleasure, and Aristotle , claims objective goods come from considering the characteristic activities of a species (such as reasoning for humans). *Desire-fulfilment* theory says welfare is the result of achieving one's desires, and *mental state theory* says welfare is merely some pleasurable mental state such as happiness. *Hedonist* theories say welfare is a matter of experiencing pleasure and avoiding pain.

Each theory has its own problems. Objective lists run the risk of including items that some individual doesn't seem to require for welfare, or missing an item that someone *does* require for welfare. Under mental state theory, a person trapped in Nozick's experience machine or the Matrix would feel (and mentally

experience) high welfare, but would (according to critics) in fact be living a miserable existence. Desire fulfilment theory implies that someone who desires to eat gravel and succeeds is in fact increasing their welfare, regardless of how irrational or unhealthy this desire is. People are frequently mistaken about their desires, and have wants which turn out to be highly undesirable once satisfied. Hedonist theories have trouble explaining heroin addiction, which appears to be a highly pleasurable yet welfare-reducing experience.

No objection is a killer knock-down argument; each theory is taken seriously by the philosophical community. This paper puts forward an objective list theory of welfare, simply because I feel such theories are best suited to analysing non-human welfare. Not all beings will have desires, pleasurable feelings or mental states, so I feel the corresponding welfare theories are ill-suited to the task at hand. This paper's primary goal is to defend its own theory of non-human welfare; not to criticise other theories of welfare. As such, my analysis of desire, hedonist and mental state theories will be limited.

My theory of non-human welfare is an extension of Daniel Haybron's self-fulfilment welfare. In (Haybron, 2008, chapter 9), Haybron proposes a eudaimonistic account of welfare which identifies self-fulfilment as the objective welfare good — in particular, fulfilment of one's emotional nature.

This paper is not a defense of Haybron; rather, I take Haybron as a suitable framework to ground non-human welfare in. Later I will defend my theory against criticism, but I will only engage with criticism of Haybron where it is relevant to my non-human welfare theory.

2   Haybron's theory of welfare

The question of welfare is what increases a life's prudential value. According to Haybron (2008, pg. 192), "some sort of nature-fulfillment is intrinsically valuable." Fulfilling one's nature increases one's welfare. But what exactly is 'nature-fulfilment'? Answering this question requires a brief explanation of Haybron's notion of the self.

By "self" Haybron means "those aspects of us that are important to making us the distinct individuals we are, that are important to understanding who we are," such as the way a person "sees or thinks about her herself, her life, her ideals, her projects and commitments, and her relationships to society and other people" (Haybron, 2008, pg. 183, 184). One's self is a sub-part of one's

nature. This means all facts about one's self are facts about one's nature, but not all facts about one's nature are facts about one's self. For example, my love of fiction is a part of my self (and thus my nature), but my height is a part of my nature and *not* my self.

Haybron's theory is eudaimonistic, i.e. it claims that welfare consists of human flourishing. Specifically, Haybron claims human flourishing consists of nature-fulfilment, primarily (but not exclusively) of the self-fulfilment kind. There are many aspects to the self (for example, we have "an emotional aspect and a rational aspect" (Haybron, 2008, pg. 193)), and when we act in accordance with some aspect, we are fulfilling our self.

Haybron is particularly concerned with our emotional nature. "To have a certain emotional nature is to be disposed characteristically to be happy in certain circumstances and not others," he writes (pg. 184). A person's emotional nature may change over time or be affected by social and cultural factors, but it is inherently *part of that person's self*. Under Haybron's theory, happiness has intrinsic prudential value because it results from fulfilling our emotional self. This means happiness is a sign of self-fulfilment and therefore of welfare (Haybron, 2008, pg. 178).

This graphic summarises the Haybron model of welfare:

[TODO: make a picture]

This account of eudaimonia differs from the traditional Aristotelian account in that it is neither perfectionist nor externalist. Both views posit that welfare requires fulfilling one's nature, but on Aristotle's view this consists in "the proper exercise of our distinctively human capacities" whereas Haybron's consists in fulfilling the "arbitrarily idiosyncratic make-up of the individual" (Haybron, 2008, pg. 193).

2.1   Welfare beyond happiness

As hinted by his book's title ("The Pursuit of Unhappiness"), Haybron's account of welfare is centered around happiness. However, Haybron explicitly states that welfare consists of more than just happiness and the fulfilment of one's emotional nature. Where else does Haybron think welfare can come from?

Haybron engages only briefly with non-happiness welfare,[1] writing that "a full-blooded account of welfare would likely incorporate goods other than hap-

---

[1] Presumably because Haybron's book is about happiness, and concerns welfare only insofar as it is relevant to happiness.

piness... it is highly plausible that self-fulfillment will involve, not just being happy, but success as well in relation to those commitments that define who we are and lend meaning to our lives" (Haybron, 2008, 193). On page 194, Haybron considers the idea that "wellbeing consists not just in the fulfillment of the self's two parts, but also in the fulfillment of our subpersonal, 'nutritive' and 'animal' natures: health or physical vitality and pleasure."[2]

I think Haybron's idea of extending welfare beyond happiness — to one's rational self and possible subpersonal nature — is a plausible ground for building a theory of non-human welfare. One's emotional and rational self-fulfilment have intrinsic prudential value because one's nature as a human is so heavily dependent on the corresponding two self-aspects. However, not all beings have emotional or rational selves. If a creature only has a subpersonal aspect (its nutritive/animal nature), then it makes sense for that aspect to define the creature's welfare. [TODO: EXPAND THIS. Maybe touch on comparison problems].

3   Non-humans and their welfare

My theory of non-human welfare is as follows. Like Haybron, I believe welfare consists of fulfilling the central aspects of one's nature. I use "central" in the sense of "central/peripheral" distinction: changes in something central to one's nature "constitute temporary changes *in* us," whereas changes in something peripheral to one's nature "happen *to* us" (Haybron, 2008, pg. 183).

Specifically, natures are composed of one or more aspects: emotional, rational or nutritive/animal. Each nature has central and peripheral aspects. Welfare is constituted by the fulfilment of one's nature's central aspects, whatever they may be.

This general theory of welfare yields Haybron's specific theory of human welfare as a special case. In humans our emotional selves and (parts of) our rational selves are central aspects of our nature. When we fulfil our emotional self by achieving happiness (in Haybron's sense of happiness as stable underlying mood of (and propensity for) psychic affirmation). We also gain welfare by acting in accordance with the special cherished values and relations that are intrinsic to our rational aspects, for example, observing strict vegetarianism, treating

---

[2]Remember that Haybron considers the self a specific subpart of one's nature. This is therefore a proposal that welfare can come from generalised nature-fulfilment rather than just the more specific self-fulfilment he considered earlier in the text. This leads him to comment "It may thus be strictly inaccurate to call this a "self-fulfillment" account" (pg. 194).

others with kindness, learning and developing as rational agents, etc.

I will now consider some examples of welfare in non-humans to demonstrate the consequences of this theory.

3.1   Animals and aliens

It is difficult to assess the intelligence of different animal species, however, some animals clearly possess a rudimentary rational mind (dolphins and great apes) while others clearly do not (worms, flies). While neither dolphins nor apes have moral obligations, values or principles to the extent that humans do, they nonetheless place value on their family members, on friendship, on acquisition of resources, etc. They seem to have emotional states such as fear and joy; they enjoy play and socialising.

This motivates my claim that some animals resemble humans in that they have emotional and rational selves which form the basis for their welfare. This seems intuitively reasonable. Apes mourn when their children die — intuitively their welfare diminishes as a result of this. Intuitively, two isolated dolphins who discover each other and begin to play are experiencing an increase in welfare.

Simpler creatures like worms or flies lack a self at all. To these creatures, I suggest welfare consists simply of their nutritive/animal nature, by which I mean their health. A worm is intuitively better-off when whole and healthy rather than when being pecked by a chicken. A fly is clearly worse-off when sprayed with Mortein.

Most animals fall somewhere between these two extremes, with some having more-developed or less-developed rational and emotional aspects than others. Research has shown that octopuses have highly-developed rational selves[3] but it is unknown how developed their emotional selves are. Creatures falling somewhere in this spectrum of self will have similarly varying welfare functions: some weighted more towards nutritive/animal health, some weighted more towards rational or emotional fulfilment.

The possible existence of aliens complicates this picture somewhat. Presumably if they exist, aliens will also have some level of self — even if, like worms, this level is zero. It could be incredibly easy to determine an alien's nature (and therefore welfare) if these aliens resembles Earth life (like the aliens in, say, Star Wars) but very difficult if we have no way to conceptualise the aliens (such as

---

[3]They are known to plan complicated procedures whereby they escape their tanks, evade aquarium staff and travel to elsewhere in the building to steal food.

in Stanislaw Lem's novel Solaris). It is possible that we will have no way to determine the relevant characteristics of an alien species, or to determine its level of welfare. But this is a purely epistemic problem, not an ontological or analytic problem. After all, the alien may have no way of comprehending human characteristics or welfare, but human welfare is still well-defined and meaningful.

3.2 Robots

Consider a stereotypical science-fiction robot: intelligent, but cold and emotionless. These hypothetical creatures have highly-developed rational self-aspects, but no emotional self-aspect.

I propose welfare for these robots consists of fulfilling their rational self-aspect. This can be justified by analogy to a human named Yvette who, by some rare condition, has been unable to feel emotions from the day she was born. You could justifiably claim she has very low welfare on the basis of being unable to feel happiness. Despite this, she still experiences welfare gains and losses thanks to her rational aspect. Yvette's welfare will decrease if her parents kick her out of home. Her welfare will increase if she finds a rewarding law job which lets her exercise her reasoning skills and allows her to leave her boring job as a retail worker.

Robots of the sort mentioned above would experience welfare in a similar way — on the basis of pure deliberation and reason, rather than emotion. This aligns with Aristotle's notion of the good life as one spent developing one's mind and reason, although without the perfectionist and externalist qualities of real Aristotelian welfare theory.

Simple robots with only minimal intelligence (for example, a Roomba autonomous vacuum cleaner) would have neither rational selves nor emotional selves, but would still have health. The labels of 'nutritive' and 'animal' don't quite fit when applied to a robot, but the general concept of health (as the conditions necessary for continued existence and functioning) still applies. Health should be considered the basis for such a creature — there is clearly some sense in which a Roomba is worse off when its circuitry becomes unreliable after extended use, or when it suffers water damage, or when one of its wheels break. Repairing these problems intuitively increases the robot's welfare. I feel justified in claiming that welfare for a simple robot like this is constituted by its 'health'.

3.3   Plants and nature

It might seem strange to consider the welfare of a plant, but governments often have to consider environmental welfare when determining policy. Deforestation and coral reef bleaching both intuitively harm the natural world, but are sometimes justified (rightly or wrongly) in the name of human welfare.

Plants obviously have no self, but they are alive and could plausibly have natures.Here, Aristotelian logic serves as a helpful guide: a plant's characteristic activities are to grow and to reproduce. If a plant's nature consists of anything it must relate to these activities.

I suggest that plants have nutritive natures (i.e. health), and that welfare for a plant consists only of this health. Intuitively, a tree is worse-off when uprooted, better off when watered appropriately, and worse-off when over- or under-watered. The tree's welfare increases with its ability to grow and reproduce. Like Haybron's original welfare theory, we can view welfare as similar to Aristotle in that it concerns the subject's characteristics, but unlike Aristotelian welfare we make no reference to the tree's perfection or external telos.

3.4   Group agents

[TODO: complete this section if you need to kill more words]

4   Evaluation

I have now outlined my account of non-human welfare and will now analyse whether it presents a useful or appealing theory. I attempt to pre-emptively address concerns readers may have.

My theory's first weakness is that it makes interpersonal comparisons of welfare very difficult. This is problematic because, as mentioned in the introduction, theories of welfare are motivated partly by the need to make decisions which affect other beings. If we cannot compare the relative welfare of two beings, or of the same being in different situations, then welfare becomes less useful as a tool for guiding decision-making.

Fortunately for my theory (and unfortunately for policy experts and philosophers) I believe any theory of non-human welfare will encounter this problem — as will many human-specific theories. It is already very difficult to compare the welfare of any two human beings in general. Yes, in some cases such com-

parisons are simple — for example, if two people have largely identical physical and emotional conditions, and then the first is waterboarded, clearly they will now have lower welfare than the second. But comparisons of welfare are not always so simple. If I spend my life working a pleasant but unrewarding job as a cashier in a rural town, and my sister moves to New York to pursue a stressful career as a lawyer, how can our relative levels of welfare be compared?

Haybron's original theory already opens itself up to indeterminacy in welfare comparisons: "some improvements in happiness, even authentic ones, may not make our lives go better for us on the whole. For they may deprive our lives of too much meaning" (Haybron, 2008, pg. 194).

References

Gert, B. (1998). *Morality: Its nature and justification*. Oxford University Press on Demand.

Haybron, D. M. (2008). *The pursuit of unhappiness: The elusive psychology of well-being*. Oxford University Press on Demand.

Joachim, H. H., & Rees, D. A. (1951). The Nichomachean Ethics (Aristotle).