# The Self-Esteem Theorem

In a number of normative domains, there are attractive views about how to evaluate our options against each other which are *decision-dependent* – where some crucial input in the evaluation may be affected or determined by the decision itself, either epistemically or metaphysically. I focus on four examples of important decision-dependent inputs from various literatures on norms of action: a) In decision theory, the likelihood of states of the world. b) In person-affecting approaches to population ethics, the identity of people who will actually exist. c) In views on which we have reasons of partial concern for with whom we have special relationships, the identity of our spouses, friends, or children. d) In views on which we have reason to satisfy our future values or preferences, the nature of those values (Laurie Paul's transformative experiences provide a vivid example).

One natural way to approach such cases is to begin with several evaluations, one from the perspective of each potential course of action, and generate an overall evaluation of the options as some function of these decision-relative evaluations. This makes the decision a kind of voting problem, where the different options play the role of both candidates and voters. We can call this the decision-relative approach.

In several of the areas above, philosophers have developed views explicitly or implicitly in line with this approach, for example, Ralph Wedgwood's benchmark decision theory. These views have tended to face objections related to a principle called Independence of Irrelevant Alternatives (IIA) – the idea that the relative evaluation of two options should not be affected on the presence or absence of a third.

In this paper, I will prove a result I call the Self-Esteem Theorem, which tells us, in effect, that there is exactly one way of generating an overall evaluation from decision-relative evaluations that does not violate IIA – each option's overall evaluation must be a function of its *self-esteem* – that is, its evaluation from its own perspective.

This means that for each method of evaluation, there will be a single decision-relative view, what we might call the *self-esteem view*, which uniquely satisfies IIA (and some other extremely minimal constraints). This view is already familiar in other guises in many of these domains – in decision theory, for example, the self-esteem view is just Evidential Decision Theory. I will discuss what the self-esteem view looks like in each of the four applications above, and why it is often resisted.

There are also, I will suggest, illuminating patterns among possible *alternatives* to the self-esteem view. Among the views which incorporate decision-relative evaluations (and therefore violate IIA), there will be an analogue *benchmark view*, for example. Some of these views exist in all of these literatures already – others represent unexplored alternatives in at least some. In each domain, for example, one possible view will be what we might call the *present perspective view*, which asks us to ignore the decision-relative evaluations in favoured of a single privileged evaluation from the perspective of the agent prior to their decision. Another will be the *actual perspective view*, which assesses options from the perspective of the action that *actually will* be chosen. Because of the structural similarities between these views, moreover, we can see in advance what potential counterexamples will look like.

I will go on to discuss what lessons we should ultimately draw from the self-esteem theorem and the rest of our investigation. One obvious possible lesson is that we ought to accept the self-esteem view in every domain, given its special advantages. But there are reasons to think some incarnations of the self-esteem view are quite unattractive. A weaker lesson is that there is some pressure on us to accept similar views across different domains – it would be a bit odd, for example, to accept the self-esteem view in one area but not the others, given that there are similar strengths and weaknesses.

Lastly, if we are (as I am) unattracted to the self-esteem view in some of these domains, but we want to take the phenomenon of decision-dependence seriously, we should look carefully at the

Independence of Irrelevant Alternatives. While every decision-relative view that is not the self-esteem view will violate it, not all of them must violate it in equally gruesome ways. I will briefly discuss weakened constraints related to IIA which the self-esteem theorem doesn't undermine. We might be okay, for example, with violations of IIA as long as introducing *dominated* options, or *impermissible* options, does not affect the relative ranking of existing options. Prominent existing views, I will show, do not satisfy these weakened constraints. But given that the self-esteem theorem shows us that a huge class of attractive potential views will have to give up IIA, exploring how much we can minimize the damage is a worthwhile project.

At the very least, this paper provides a framework for looking at a set of debates which have taken place more or less independently in different normative literatures, and which allows us to draw parallels between views and problems in each. The self-esteem theorem is just one result with a broad application across debates where decision-dependence arises. I expect that there are many more waiting to be discovered.