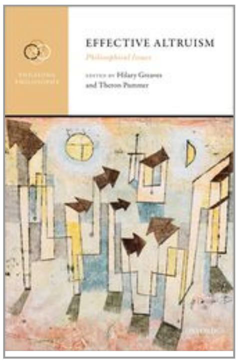

Effective Altruism: Philosophical Issues
Hilary Greaves and Theron Pummer

Print publication date: 2019
Print ISBN-13: 9780198841364
Published to Oxford Scholarship Online: October 2019
DOI: 10.1093/oso/9780198841364.001.0001

Effective Altruism and Transformative Experience

Jeff Sebo
Laurie Paul

DOI:10.1093/oso/9780198841364.003.0004

Abstract and Keywords

In this chapter, Jeff Sebo and L.A. Paul investigate the phenomenon of experiences that transform the experiencer, either epistemically, personally, or both. The possibility of such experiences, Sebo and Paul argue, frequently complicates the practice of rational decision-making. First, in transformative cases in which your own experience is a relevant part of the outcome to be evaluated, one cannot make well-evidenced predictions of the value of the outcome at the time of decision. Second, in cases in which one foresees that one’s preferences would change following the decision, there are issues about whether rational decision-making should be based only on one’s ex ante preferences, or should also incorporate some element of deference to foreseen future preferences. While these issues arise quite generally, Paul and Sebo suggest that they are especially pressing in the context of effective altruism.

Keywords: transformative experience, cost–benefit analysis, peer disagreement, decision theory, collective agency, authenticity, narrative self

1. Introduction

Effective altruists try to use evidence and reason to do the most good possible. However, some choices involve transformative experiences, which change what we care about in ways that we cannot fully anticipate. This limits our ability to make informed, rational, and authentic plans individually as well as collectively.

In this chapter, we discuss the challenges that transformative experiences pose for effective altruists, given that such choices change us in surprising ways.

2. Effective altruism

Many effective altruists think about what to do in the following kind of way: First, they think about the *scale* of a problem. The more harm a problem causes, the higher priority it should have according to effective altruism all else equal. Second, they think about how *neglected* a problem is. The more neglected a problem is, the higher priority it should have according to effective altruism all else equal. Third, they think about the *tractability* of a problem. The more tractable a problem is, the higher priority it should have according to effective altruism all else equal. Finally, they think about *personal fit*. Given everything they know about their talents, interests, and backgrounds, what can they do individually in order to address the worst, most neglected, most tractable problems as effectively as possible?[1]

Many effective altruists try to answer these questions through impartial cost–benefit analysis. They try to collect as much evidence as possible, assign probabilities and utilities to different courses of action on the basis of this evidence, and then select the course of action that maximizes expected utility. Moreover, many effective altruists do not assign special weight to what they, as individuals, happen to think or feel. Yes, they care about personal fit, but only from an impartial standpoint. **(p.54)** They think that they should do the most good possible for everyone in the world, and so personal fit is relevant primarily insofar as it impacts productivity. Similarly, they care about deliberating about which course of action is best, but, again, only from an impartial standpoint. They think that they are only one of many people asking these questions, and that if they disagree with other, seemingly equally informed and rational individuals about the answers, they should seriously consider the possibility that they are wrong.

Given this commitment to informed, rational, impartial benevolence, effective altruists tend to agree about many issues. For example, they tend to agree that existential risk, global health and development, and animal welfare are high-priority cause areas.[2] They also tend to agree that certain interventions in these areas are more effective than others. Within the animal welfare category, for example, they agree that farmed animal advocacy is a higher priority than companion animal advocacy.[3]

With that said, effective altruists also disagree about some issues. For example, they disagree about some normative issues, such as whether one should attempt to maximize happiness or merely minimize suffering, and about whether one should do so by any means necessary or while respecting deontological side constraints. They also disagree about some descriptive issues, such as what kind of effective altruist movement is likely to produce the relevant desired outcomes, or what kind of political or economic system is likely to do so. (We will return to these issues below.) These methodological commitments, together with these areas of agreement and disagreement, raise several challenges for the effective altruist, two of which will be our focus here.

The first challenge concerns cost–benefit analysis. Effective altruists aspire to use cost–benefit analysis to decide what to do, yet they often lack essential information. In this kind of case, should they still attempt to apply cost–benefit analysis to all relevant options? Or should they apply cost–benefit analysis to a narrower range of options and/or use a different decision procedure?

The second, related challenge concerns impartiality. Effective altruists aspire to reason impartially, yet they do not always reach the same conclusions as other, seemingly equally informed and rational individuals. In this kind of case, should they assign weight only to the beliefs and values that they identify with, or should they assign weight also to other, seemingly equally informed and rational beliefs and values that they feel alienated from?

In what follows, we will explore how the possibility of undergoing a transformative experience can exacerbate these challenges for effective altruists, individually and collectively.

(p.55) 3. Transformative experience

An experience can be transformative in at least two related ways. First, an experience is *epistemically transformative* when it teaches you something you could not have learned without having that experience. By having it, it teaches you what that kind of experience is like, and it also gives you the ability to imagine, recognize, and cognitively model new possible states. For example, you can learn what parenthood is like for you only by actually becoming a parent.[4] Second, an experience is *personally transformative* when it changes you in a personally fundamental way by changing a core personal belief, value, or practice.[5] For example, by becoming a parent, you can acquire an updated set of beliefs, values, and motivations. There can also be a certain amount of endogeneity. For instance, many parents find that, after having a child, they form a preference to have had that very child. In light of such changes, your pre-decision (*ex ante*) self and your post-decision (*ex post*) self might have different preferences, including different higher-order preferences.[6] A *transformative experience*, as defined by Paul, is an experience that is both epistemically and personally transformative.[7]

There are many ordinary examples of transformative experience. Some are relatively sudden, such as the experience of moving to a new city, starting college, starting a new job, having a baby, experiencing violent combat, or gaining a sensory ability. Others are gradual, such as the transformation from being ten years-old to being thirty years-old, from being a graduate student to being a tenured professor, or from being a Syrian refugee to being a U.S. citizen.

Either way, these transformations are all in a certain sense irreversible. You can drop out of college, leave your job, and even leave your family, but these experiences will have affected you (in addition to having opportunity costs and changing your choice situation).

When a person thinks about what to do, they have to consider many possible things they could do, but in transformative contexts, they must also consider the many possible selves they could become. When these changes will be irreversible, a person has to decide what to do without having the opportunity to experience these different futures. So, if a person is making a decision that may involve transformative experience, they have to decide what to do without knowing what it will be like to take each available path. They also have to decide what to do even if this decision could change their core beliefs or values in a way that creates *ex ante/ex post* conflict.

(p.56) The possibility of transformative experience exacerbates the challenges for effective altruists that were considered in the previous section. First, it exacerbates the challenge to cost–benefit analysis, by raising the question of how to decide what to do if you will learn essential information only after the decision is made. For example, if you can accurately imagine parenthood only after becoming a parent, how do you decide whether or not to become a parent?

Importantly, the challenge is not merely that, prior to making your choice, you are uncertain about the probabilities and utilities of the outcome. The challenge is also that *you cannot assign value to the outcome with any accuracy*. Your value function for the outcome is undefined. This is because you cannot imaginatively represent an essential part of the outcome (the nature of the lived experience of being a parent) well enough to accurately assess its value.

Why, exactly, does your value function for the outcome go undefined? For the familiar reason that the relevant information carried by the experience cannot be grasped without having the experience. It is not possible, for example, for a person who has never seen color to know or accurately imagine what it is like to see red. She needs to have the experience before she can assign value to what the experience is like (at least, with any accuracy). Other transformative experiences are similar. In each case, we cannot know or accurately imagine what it is like to have a fundamentally new kind of experience until we have actually had it. And, insofar as we need to assign value to what the experience is like in order to assign value to an outcome involving that experience, our inability to make the former assignment with any accuracy will lead to an inability to make the latter assignment with any accuracy.[8]

This is therefore more than a case of uncertainty: It is a case of ignorance. And in many cases, this ignorance will never be fully resolved, not even after the fact. If you make one choice, you will bring about one future as a result, which you will then be able to accurately value and represent. But not only will you have already made your choice at this point, you will also still be unable to accurately value the other futures that you could have brought about through other choices. Therefore, you will still be unable to assess your choice relative to other choices that you could have made. The question, then, is: How should you decide what to do? Should you use cost–benefit analysis and consider all relevant options, even if you are unable to assess them? Or should you consider only options you are able to assess, or use a different decision procedure?[9]

Of course, to say that we lack essential information for first-person value assessment is not to say that cost–benefit analysis is always useless. Some cases **(p.57)** are relatively easy to resolve without first-person value assessment, since they involve changes that are always good (or bad). Other cases are harder to assess, but we might still have at least some evidence to draw from, such as evidence about how other people react to this kind of change or how we react to other kinds of change. Alternatively, we might lack evidence but still have speculative estimates to draw from.[10]

However, it is not clear that these considerations will be enough to make cost–benefit analysis useful in the kinds of cases that we are discussing here. First, even when we do have evidence, it is not clear how representative this evidence is. Seeing how other people react to this kind of change will not necessarily tell us how *we* will react to it, and seeing how we react to other kinds of change will not necessarily tell us how we will react to *this* kind of change. Second, while speculative estimates can often be useful, it is not clear that they can be useful in many transformative cases, since, as noted above, we cannot assign value to all outcomes before having the experience and our preferences may be endogenous.

The possibility of transformative experience also exacerbates the challenge to impartiality, by raising a question about how to make decisions in cases where your core personal beliefs and values might change as a result. For example, if your preference for being a parent is endogenous to the process of becoming a parent, should you base your decision about whether or not to become a parent on an evaluative standpoint that excludes or includes this preference?[11] Moreover, if we suppose that you should do the latter, what happens if you expect to have *ex ante/ex post* conflicts arise? For example, what if you currently have one preference (e.g. to have one child), but you expect to form another if you end up remaining a non-parent (e.g. in the future you expect to prefer to have no children). What if you prefer to have one child now, but you expect to prefer to have twins (triplets ….) if you end up having twins (or triplets)?

There are other reasons why one might care about the prospect of preference change. Some are, appropriately, existential in nature. For instance, you might resist making decisions that, in your view, would result in an elimination of your current self. Similarly, if you care about first-personal deliberation, then you might resist basing your decisions in part on preferences that you currently feel alienated from. But since many effective altruists care more about doing the most good possible than about avoiding self-elimination or alienation, we will not focus on that issue here.[12]

Other reasons for caring about the prospect of preference change are prudential, moral, or political in nature. For example, if you think that you have prudential, moral, or political duties to your future selves, then you might think that you **(p.58)** should allow them to have a say in your decision as a matter of prudence, morality, or justice.[13] However, since effective altruists tend to care more about doing the most good overall than doing the most good for themselves (except insofar as they think that the ability to compromise and coordinate with past and future selves is instrumentally valuable), we will once again focus on other issues.

Importantly, groups may be able to have transformative experiences as well. Groups may not have phenomenally conscious mental states in the same kind of way that individuals do, but they can still have beliefs, values, and preferences in the relevant sense. For example, they can construct these states directly, by endorsing certain statements of fact, value, and priority. They can also construct these states indirectly, by pursuing courses of action that make sense in light of certain belief, value, and priority attributions. Either way, as in the individual case, groups will tend to form beliefs, values, and preferences that make sense in light of their actions and will tend to perform actions that make sense in light of their beliefs, values, and preferences. Moreover, as in the individual case, group members might sometimes face decisions that could change the group in ways that are difficult to anticipate, and which could result in *ex ante/ex post* conflict. For instance, if a company hires a new staff member or implements a new policy, they need to consider the possibility that this decision will result in preference change for the company as a whole.[14]

As in the individual case, the possibility of transformative experience exacerbates the challenges considered above. For example, when a company has to make a decision that may result in a transformative experience, should they use cost–benefit analysis and consider all relevant options, or should they consider fewer options and/or use a different decision procedure? Also, should they base decisions entirely on their current beliefs and values, or should they defer at least partly to other beliefs and values? Once again, one might care about these questions for many reasons. But we will here focus on the reasons for which an effective altruist will care about them.

Whether we confront cases involving transformative experience individually or collectively, we face the following kind of tension: Insofar as we restrict what we think about and how we think about such cases, we will be able to reason relatively accurately and authentically, but we will also limit our opportunities for doing good. Whereas insofar as we expand what and how we think about such cases, we will be able to consider more opportunities for doing good, but we will also recognize new limitations on our ability to reason accurately and authentically.

In what follows we will consider some examples that illustrate the challenges that choices involving transformative experiences raise for effective altruists. **(p.59)** We will explore these challenges at both the individual and collective level, showing that analogous challenges arise at both levels, and suggesting that the stance effective altruists take toward such challenges will have a pervasive influence on their decision-making and impact.

4. Individual transformation

Effective altruists, like anyone else, face transformative choices such as what to do for a living, whether to get married, whether to have kids, and so on. Managing such choices can be especially challenging for an effective altruist, since in each case they are committed to using evidence and reason to do the most good possible, which requires deep assessment of a wide range of options. We will here focus on career choice as an illustration, but similar questions will arise for other choice situations as well.

Suppose that you are an effective altruist deciding what to do for a living,[15] and that you have three main options to consider: You can (*a*) go to grad school (so that you can work in research and education), (*b*) go to law school (so that you can work in law and politics), or (*c*) work in finance (so that you can earn to give). Suppose also, since grad school and law school would be more continuous with your college experience than finance would be, you have a better sense of what your life would be like in the first two scenarios than in the third.

In particular, the choice whether to work in finance strikes you as high risk/high reward. If it works out, you could earn millions of dollars per year and then donate that money to effective causes. But you wonder if you can expect it to work out. Here you may ask: Would I fail at investment banking? Would I succeed but lose my commitment to effective altruism? Would I retain my commitment to effective altruism but start to think that I need to spend more money on myself than I currently think I do? If I did change my mind in one or more of these ways, would I be rationally updating in light of new information and arguments? Would I simply be rationalizing the kind of self-interested behavior that I would have, at that point, been socialized into? Or might I change in other ways that I cannot imaginatively anticipate, and which might raise other possibilities for *ex ante/ex post* conflict?

With this in mind, consider the challenge that this kind of transformative choice can raise for cost–benefit analysis. For some people, the costs and benefits of these options might be easy to assess. For example, if you find that you have very little interest in material things and that your social environment has very little impact on your beliefs and values, then it might be rational for you to feel **(p.60)** confident that working in finance is the right choice for you. Likewise, if you find that you have a lot of interest in material things and/or that your social environment has a lot of impact on your beliefs and values, then it might be rational for you to feel confident that working in finance would be wrong for you. (Though even in these cases mistakes are possible.)

But for others, the costs and benefits of these options might be harder to assess. For example, if you find that you have a decent amount of interest in material things and/or that your social environment has a decent amount of impact on your personality, then it might not be rational for you to have much confidence one way or the other about whether finance would be right for you. For all you know now, if you worked in finance, you could be happy, productive, and committed to effective altruism and to earning to give. Or you could be happy, productive, and uncommitted. Or you could be miserable, productive, and committed. Or you could be miserable, unproductive, and uncommitted. And so on.

If you find yourself with this kind of question, how should you go about making this choice? A natural thought is to apply cost–benefit analysis to all of your options to the best of your ability. You can collect as much evidence as possible and then make the choice that maximizes expected value, given your evidence. In this case you have to ask: What kind of evidence is available to me?

One source of evidence comes from other people in this situation. Now that more people are earning to give, more information is available about successes and failures. But insofar as an effective altruist is interested in evidence-based estimates of value (as opposed to speculative estimates of value), what matters is not information in the form of anecdote, unvetted testimony, or emotional appeal. Rather, what matters is evidence drawn from long-term, empirically rigorous case studies. A problem here is that, since the effective altruism movement is fairly young, such evidence is not yet available.[16] Moreover, even if you were to have access to evidence from long-term, empirically rigorous case studies on other people, that might still not be enough to tell you what it will be like for *you* to be in this situation. As with any complex life experience, there is enough heterogeneity amongst individuals to raise worries about your ability to discover your reference class. That is, you need to know whether you are relevantly similar to other effective altruists to know whether working in finance would have the same impact on you as it did on those for whom data is available.

A second source of evidence comes from you in other situations. You might not have the experience of taking on the role of investment banker, but you have experience taking on other social roles, and then observing whether and to what degree these choices affect you. Perhaps in the past you remained happy, productive, and committed to effective altruism in the face of changing social environments. **(p.61)** But once again, what matters is not information in the form of your own memory and self-narrativity, but *evidence*. You need evidence that rules out the possibility that there are relevant differences between this situation and other situations, differences that are opaque to you now, in virtue of which this choice would have a different impact on you than other choices did.

A third, related source of evidence is what John Stuart Mill called experiments in living.[17] You can dip your toes in the water by taking classes in finance, taking a summer internship in finance, spending time with people who work in finance, and so on, and, as a result you can collect evidence about yourself in this situation without yet committing to this path. This can certainly help. But insofar as these experiments are informative, they may also be transformative: You may already be changing your preferences as a result of the experience. And, insofar as these experiments are not transformative, they may also not be informative: You may still be making a decision about what to do in a state of ignorance about what it will be like to fully take this path.

Note that with respect to all three sources of evidence (especially the latter two), there is a risk of confabulation and cognitive dissonance that you will also need to address, insofar as you were committed to using evidence over anecdote, testimony, or hope when making important choices. There is also a risk that, if you have more familiarity with some options than with others, then your application of cost–benefit analysis will reflect bias. In some cases, this might mean a bias in favor of the status quo, resulting from the availability heuristic, status quo bias, sunk cost reasoning, and so on.[18] In other cases, it might mean a bias in favor of alternatives to the status quo, resulting from selective and wishful thinking about the nature and value of unknown possible futures.

Alternatively, you can try to decide in a different way. For example, you can use cost–benefit analysis while focusing only on options that you can accurately imagine, where presumably this means going to grad school or law school. You can err on the side of caution, where again presumably this means going to grad school or law school. You can do what makes you happy in the moment. You can make a radical choice, where this could mean any number of things. And so on.

To be clear, these decision procedures can be justified within an effective altruism framework. If evidence and reason indicate that you can do more good by using an alternative to cost–benefit analysis in some cases than by using cost–benefit analysis in all cases, then cost–benefit analysis at the meta level can endorse alternatives to cost–benefit analysis in some cases at the object level. If you reach this conclusion, then you would be a kind of indirect effective altruist, similar to indirect utilitarians who think that utilitarianism at the meta level endorses alternatives to utilitarianism in some cases at the object level.

(p.62) A challenge for this indirect approach, however, is that in cases involving transformative experience, you lack information about not only which *choice* will be best but also which *decision procedure* will be best. Granted, as above, you can ask what decision procedures tend to be useful for others in this kind of situation and for you in other kinds of situation. But you would still face the same challenges, only at a higher level. The kind of evidence you would need is difficult to collect. Moreover, evidence about which decision procedures work for others in this kind of situation will not tell you which decision procedure will work for you in this kind of situation, and evidence about which decision procedures work for you in other kinds of situation will not tell you which decision procedure will work for you in this kind of situation. And, insofar as this is true, you will once again be at risk of bias if you try to use intuition, speculation, anecdote, and so on to fill in the blanks.

Consider now the challenge that this kind of transformative experience can raise for impartiality. How should you go about making this choice if it might produce *ex ante/ex post* conflict? That is, how should you decide what to do if there is a reasonable chance that becoming an academic, a lawyer, or an investment banker will give you preferences that differ from your current preferences?[19] Should you base your decision entirely on your current preferences, or should you defer at least partly to your expected future preferences? Moreover, if there is no perspective-independent, higher-order way to resolve these differences, how can such a choice be rational?

One option is to endorse *the ex ante privilege view* and act only on the basis of your current preferences. On this view, you can consider the possibility of a change in preferences, but only to inform your current perspective. For example, if you expect your preferences to change, you can ask if your future self has preferences that your present self prefers (and, if so, you can update your current preferences accordingly). Similarly, you can reflect on how this change in preferences could be a problem for your current plans (and, if so, you can update your current plans accordingly). But beyond that, you should not, on this view, consider assigning any independent weight to your expected future preferences. For example, you should not think, “I reject my expected future preferences, and I do not see them as a threat to my current plans. But I will defer partly to them anyway.” The benefit of the *ex ante* privilege view is that it coheres with standard decision theory, makes your deliberation relatively simple, and allows you to act only on preferences that you currently identify with. However, the cost of this view is that it arguably conflicts with the kind of impartiality that many effective altruists aspire to. After all, if you expect to have different preferences in the future from the ones you have now, and if you expect to be at least as informed and rational in the **(p.63)** future as you are now, then why does it make sense for you to privilege your current preferences over your expected future preferences when deciding what to do?

Another option, then, is to accept *the equal weight view* and act on the basis of an evaluative perspective that assigns equal weight to your current preferences and your expected future preferences. As with alternatives to cost–benefit analysis, this view can be justified within an effective altruism framework. In particular, if evidence and reason indicate that you can do more good by assigning weight to multiple, conflicting perspectives, then your current, pro-effective altruist preferences at the meta level can endorse this approach at the object level. Moreover, the equal weight view is arguably more impartial than the *ex ante* privilege view, since, again, it seems arbitrary for you to privilege your current preferences over later, equally legitimate preferences. At the same time, the equal weight view departs from standard decision theory, makes your deliberation more complicated, and may require you to act at least partly on preferences that you currently feel alienated from (including, possibly, non-effective altruist preferences).[20]

If you think that you should assign at least some weight to your expected future preferences, then you face additional questions about how extensive your epistemic humility should be. Consider three such questions as an illustration.

The first question concerns the distinction between actual and possible preferences. Recall that each choice you can make would bring about a different set of beliefs and values. This raises the question: Should you consider only the preferences that you would *actually* have, given the relevant choice, or should you consider also any preferences that you could *possibly* have, independently of the relevant choice? On the former, narrow view, you would consider your expected academic preferences when considering becoming an academic, your expected lawyer preferences when considering becoming a lawyer, and your expected investment banker preferences when considering becoming an investment banker. Whereas on the latter, wide view, you would consider all three sets of preferences when considering all three choices.

The narrow view simplifies deliberation, but it can also lead to bias. After all, why should you think that the preferences that you would have given your actual choices are more likely to be informed or rational than the preferences that you would have given other, possible choices? The narrow view can also lead to problems. For example, what should you do if your expected academic self prefers that you become a lawyer, your expected lawyer self prefers that you become an investment banker, and your expected investment banker self prefers that you become an academic? Meanwhile, the wide view avoids bias and paradox, but it complicates your deliberation.

(p.64) The second question concerns intrapersonal versus interpersonal preferences. In particular, should you consider only the preferences that *you* would or could have, or should you consider also the preferences that *others* would or could have? On the former, narrow view, you would consider your actual and possible future preferences, but not others'. Whereas on the latter, wide view, you would consider others' as well. So, for example, even if you could never be from a different nation or generation, you might still have reason to consider the actual or possible preferences of individuals from other nations or generations when deciding what to do.

As with the previous issue, the narrow view simplifies deliberation. It also affirms the importance of personal identity, since it implies that you have reason to assign weight to your own preferences as such (assuming that you care about that). However, the narrow view can also lead to bias and paradox. Meanwhile, the wide view avoids bias and paradox, but it also complicates your deliberation. Granted, you might still have reason to assign extra weight to your own preferences in practice, since many of your plans will require cooperation from your future self. But this is a practical, not a theoretical, consideration, and it might apply more in some cases than in others.

The third question concerns how to determine whose preferences you should not only *consider* assigning weight to but *actually* assign weight to. Should you make use of your current beliefs and values while making this determination, or should you bracket them while doing so? On the former, narrow view, you might determine whose preferences count in part by asking if they share your commitment to effective altruism. On the latter, wide view, you would have to determine whose preferences count independently of whether or not they share this commitment, and so you would likely end up assigning weight to a wider range of preferences.

As before, the narrow view simplifies deliberation, in part by giving you a relatively clear basis for determining who counts. After all, if you bracket all your beliefs and values when stating and evaluating different sets of beliefs and values, then it is unclear how you can state or evaluate them at all. However, the narrow view can also lead to bias. After all, insofar as you require others to share your beliefs and values in order to count, you risk biasing your deliberation. Meanwhile, the wide view will not lead to bias, but it will also complicate your deliberation by raising questions about how you can evaluate other preferences without access to any particular standard of evaluation.

We cannot fully evaluate these issues here. However, we will note that, since narrow and wide answers to these questions tend to have similar pros and cons, we have at least some reason to expect that an effective altruist should take either a narrow or wide approach across the board. Moreover, since effective altruists tend to favor impartiality over partiality, we have at least some reason to think that effective altruists will tend to favor a wide approach to a narrow approach across the board. If this is right, then there is no special question about whether **(p.65)** you should, say, assign weight to your expected investment banker preferences. Instead, the question is a much more general one: whether you should assign weight only to your own current preferences or also to many other preferences, actual and possible, future self and other, friend and enemy (where many effective altruists will likely prefer the latter option, perhaps within certain limits).

As this discussion shows, how we approach these problems involving cost–benefit analysis, impartiality, and transformative experience can have a significant impact on our decision-making, and there is no obvious or simple response. In our test case, insofar as you take a cautious approach by restricting yourself to options you can imagine and perspectives you can endorse, you will be able to reason relatively accurately and authentically, but you will not be considering all relevant possibilities. For instance, you might not consider finance or your expected future preferences at all, in spite of the fact that doing so might be necessary for doing the most good possible. Whereas insofar as you proceed more adventurously by allowing for other options and preferences, you will be considering all relevant possibilities, but you might not be reasoning accurately or authentically. For instance, you might decide to pursue finance based on partial deference to expected future preferences that you can barely imagine, let alone endorse.

5. Collective transformation

As noted above, groups can have transformative experiences too, many of which will be relevant for effective altruists trying to decide what to do. These transformative experiences can occur at many levels. For example, many effective altruists live and work together in small groups. They also, in part through these small groups, participate in the effective altruism movement. And they also, in part through the effective altruism movement, contribute to society as a whole.

In each of these cases (and many others), effective altruists are part of a group that has beliefs, values, and preferences in the relevant sense. And, in each case, group members might sometimes face decisions that could change the group in ways that are difficult to anticipate, and which could result in *ex ante/ex post* conflict. For example, you might be considering adding a new roommate to your apartment or implementing a new chore system in your apartment. You might be considering hiring a new staff member at work or implementing a new division of labor at work. You might be considering advocating to expand or redistribute power within the effective altruism movement. You might be considering advocating to open borders or redistribute benefits and burdens within your society. And so on. If so, then, in each case, you need to consider the possibility that these actions will result in transformative change for the group as a whole.

(p.66) Granted, the details will vary from case to case. For example, the sense in which a household thinks and acts collectively is of course much different than the sense in which a society thinks and acts collectively. Still, insofar as these groups think and act collectively, and insofar as effective altruists can shape what these collective thoughts and actions are, effective altruists will face similar challenges with respect to these groups to those they face as individuals. In particular, in both cases, effective altruists will have many options to consider, where some of these options will be relatively continuous with the status quo and others will be relatively discontinuous with the status quo. And, the options that are relatively continuous with the status quo will be easier for effective altruists to imagine and less likely to result in fundamental change than options that are relatively discontinuous with the status quo. As a result, effective altruists will face the challenges concerning impartial cost–benefit analysis that we considered above.

First, effective altruists will have to decide whether to use cost–benefit analysis and, if so, whether to apply this framework to a narrow or wide range of options. Here they will face the same tension as before. Insofar as they apply cost–benefit analysis to a narrow range of options, they will be able to reason reliably about the options they consider, but they will not be able to consider all relevant options. Whereas, insofar as they use a different decision procedure or consider a wide range of options, they will be able to consider all relevant options, but they will not be able to reason reliably about them.

Second, effective altruists will have to decide whether to make these decisions only from the standpoint of their current preferences, or whether to at least partly defer to other preferences as well, including but not necessarily limited to their own expected future preferences. Here too they will face tension. Insofar as they consider only their current preferences, they will be able to reason authentically, but will not be able to consider all relevant preferences. Whereas insofar as they consider other preferences as well, they will be able to consider all relevant preferences, but they might not be able to reason authentically or rationally.

Of course, to say that we face similar questions in the individual and collective cases is not to say that we should answer them the same way in all cases. For example, it might be that we should take one approach in cases involving individual or small group change, and then another in cases involving medium or large group change. Still, we need to consider each case carefully. Otherwise we might find ourselves simply defaulting to a particular approach, either cautious or adventurous, without appreciating how sweeping its implications can be across cases. For example, at the cautious end of the spectrum, we might find ourselves placing strict limits on the goals that we pursue not only for ourselves but also for society as a whole, simply on the grounds that they happen to be the options we are currently able to imagine and endorse. Whereas at the adventurous end of the spectrum, we might find ourselves pursuing a deeply odd set of **(p.67)** personal and societal goals, involving outcomes that we are unable to even imagine, let alone endorse.

6. Conclusion

This chapter has sketched some of the challenges that arise for decision-making in the context of individual and collective transformation, especially as such challenges can arise for the effective altruist. We hope we have shown that the question of how to make informed and rational decisions in transformative contexts is interesting and worth further study within the context of the effective altruism movement. The effective altruist should be concerned about these problems, since, if they act without understanding or managing them, they risk either missing the possibilities that they need to consider in order to do the most good possible, or losing the focus, empirical rigor, and philosophical sophistication that makes effective altruism distinctive.[21]

References

Bibliography references:

Animal Charity Evaluators. 2018. "Donation Impact." Available at https://animalcharityevaluators.org/donation-advice/donation-impact/.

Becker, Howard S., Blanche Geer, Everett C. Hughes, and Anselm Strauss. 1961. *Boys in White: Student Culture in Medical School*. Chicago: University of Chicago Press.

Briggs. 2015. "Transformative Experience and Interpersonal Utility Comparisons." *Res Philosophica* 92 (2): 189–216.

Dennett, Daniel C. 1992. "The Self as a Center of Narrative Gravity." In *Self and Consciousness: Multiple Perspectives*, F. Kessel, P.M. Cole, and D. Johnson, eds. Hillsdale, NJ: Erlbaum.

Jackson, Fred. 1986. "What Mary Didn't Know." *The Journal of Philosophy* 83 (5): 291–5.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

MacAskill, William. 2015. *Doing Good Better*. Norwich: Guardian Faber Publishing.

Mill, John Stuart. 2004. *On Liberty*. New York, Barnes & Noble Books.

Open Philanthropy Project. 2018. “Focus Areas.” Available at https://www.openphilanthropy.org/focus.

(p.68) Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Pettigrew, Richard. 2015. “Transformative Experience and Decision Theory.” *Philosophy and Phenomenological Research* XCI: 3.

Pettigrew, Richard (forthcoming). *Choosing for Changing Selves*. Oxford: Oxford University Press.

Paul, L.A. 2014. *Transformative Experience*. Oxford: Oxford University Press.

Paul, L.A. 2015a. “What You can’t expect when you’re expecting.” *Res Philosophica* 92: *2*.

Paul, L.A. 2015b. “Transformative experience: precis and replies.” *Philosophy and Phenomenological Research* XCI (3): 760–5.

Paul, L.A. and Kieran Healy. 2017. “Transformative Treatments.” *Nous* 52 (2): 320–35.

Paul, L.A. and John Quiggin. 2018. “Real World Problems.” *Episteme* 15 (3): 363–82.

Schectman, Marya. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.

Schweikard, David, and Hans Schmid. 2013. “Collective Intentionality.” *Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.). Available at https://plato.stanford.edu/archives/sum2013/entries/collective-intentionality/.

Sebo, Jeff. 2015a. “The Just Soul.” *Journal of Value Inquiry* 49 (1–2): 131–43.

Sebo, Jeff. 2015b. “Multiplicity, Self-narrative, and Akrasia.” *Philosophical Psychology* 28 (4): 589–605.

Sebo, Jeff, and Peter Singer. 2018. “Activism.” In *Critical Terms for Animal Studies*, Lori Gruen, ed. Chicago: Chicago University Press.

Singer, Peter. 2015. *The Most Good You Can Do.* New Haven: Yale University Press.

Velleman, J. David. 2009. *How We Get Along*. Cambridge: Cambridge University Press.

Notes:

[1] MacAskill (2015); Singer (2015).

[2] Open Philanthropy Project (2018).

[3] Animal Charity Evaluators (2018).

[4] We think this problem, as it occurs in the real world, is both serious and often underestimated by philosophers. See: Paul and Quiggin (2018).

[5] Note that this sort of self-change need not entail a change in personal identity.

[6] Paul (2014); Pettigrew (2015); Paul (2015a); Paul and Healy (2017); Paul and Quiggin (2018).

[7] Paul (2014).

[8] This predicament is especially severe in real life cases, since we can't exploit the theoretical possibility that we could know what an experience is like simply by knowing, in complete detail, the neurological states that would realize that experience. For further discussion of the color vision case, see Jackson (1986). For further discussion of the parenthood case, see Paul (2014, ch. 2).

[9] For further discussion, see Pettigrew (2015); Paul (2015b).

[10] See Askell, Chapter 3 in this volume.

[11] Paul (2014); Paul (2015b).

[12] For discussion of the unimportance of the self and personal identity in prudence, morality, and rationality, see Parfit (1984).

[13] Briggs (2015); Sebo (2015a).

[14] For discussion of the idea of collective agency, see Schweikard and Schmid (2013). For discussion of the idea of collective self-narrativity, see Sebo (2015b). And, for discussion of the role of self-narrativity in self-constitution, see Dennett (1992); Schectman (1996); and Velleman (2009).

[15] For some anecdotal information about how effective altruists think about career choice, see the resources at 80,000 Hours: https://80000hours.org/

[16] For related problems with the interpretation of observational data as well as with applying such results to one's own case, see Paul and Healy (2017) and Paul (2015a).

[17] Mill (2004, p. 59).

[18] For more on cognitive biases, see Kahneman (2011). For related discussion of how these biases can be relevant to effective altruism, see Sebo and Singer (2018).

[19] For a classic description of preference change in medical students see Becker et al. (1961).

[20] There are other views as well, including the *ex post privilege view* (act in accordance with your expected future preferences). But we will focus on the *ex ante* privilege view and the equal weight view here. See Pettigrew (forthcoming) for more sophisticated treatments.

[21] Thanks to the editors of this book and to the organizers and participants of the 2017 American Philosophical Association Pacific Division Meeting, the Reading Group on Transformative Experience at UNC-Chapel Hill (especially Chris Blake-Turner), the Chapel Hill Workshop on Transformative Experience at UNC-Chapel Hill, and the Conference on the Ethics of Giving at the University of St. Andrews.