

## Effective Altruism: Memories of the Future and Simple First Principles

### Three Scenarios: Which Is Worse?

Feel into the following three scenarios. In the first, the world is at peace. In the second, 99 percent of the world is destroyed. In the third, the final one percent is destroyed. As a type of thought experiment, ask yourself which is worse: the movement from the first to the second possibility or from the second to the third possibility? Intuitively, we might think that the amount of damage in the movement from peace to 99 percent of the human world's death is obviously worse. In the move between scenario 1 and 2, some 7.7 billion people die. In the move from 2 to 3, perhaps another 78 million people die. Despite the fact that life is not reducible to quantified measurement, it seems almost self-evident that 7.7 billion deaths is far more disastrous than 78 million deaths.

But in some deep sense, this is not the case. For the death of the final one percent means also the death of the entire future. The move from 2 to 3 desecrates not only the present and the memory of the past but the memory of the future.

All of the unborn, all of the life, all of the hope, all of the goodness, truth, and beauty of billions of future years is destroyed. Reality is just arriving at a moment where there is a critical mass of human beings who stand for universal human rights, for a shared vision of goodness, truth, and beauty, for the dignity and nobility of every human being whose story deserves to be lived, shared, and celebrated, for the dignity of the animal kingdom and of the planet itself in all of its forms of life.

### A Covenant Between Generations

Human beings have been alive for some ten thousand generations. And there is an implicit evolutionary covenant between the generations. To get a sense of what this might mean let's turn for a moment to the individual and then back to the community or what we have called the generation.

Every person is responsible for their own evolutionary transformation—a continuation of the process of continuous transformation which science has revealed as the essence of evolution itself. Just like every person, every generation is responsible for its own evolution of consciousness. Each generation, building on what came before or learning from previous mistakes, commits to contribute its own unique insights and technology to the ongoing transformation and evolution of culture and consciousness. It is in that sense that there is a covenant of partnership between the generations.

But we are not an ordinary generation. We stand, as no other generation has, poised between utopia and dystopia. Our exponential tech and its democratized proliferation among state and non-state actors creates exponential risk of both catastrophic and existential nature. The social fabric of polarized discord, sans a shared global ethos for a global civilization, increases dramatically the cast of possible disaffected actors. The ostensible post-truth lack of a shared narrative of value—the good, the true, and the beautiful—combined with broken narratives of identity and community, makes it all the more dangerous. Toby Ord, in his recent book on existential risk [which appeared after Barbara passed], gives the odds of extinction of this next chunk of history at 1/6 within the relatively immediate time frame, and at 1/2 in what is, from a historical perspective, an only slightly longer time frame. We are undergoing a phase shift the likes of which the world has never seen before. This has brought us to a pivoting point at an unprecedented historical crossroads. We can evolve to ever-greater freedom, consciousness, and elegant order—a promise beyond imagination—or we can devolve into pain and peril beyond imagination.

Responding to existential risk then becomes the erotic and ethical imperative of this time. We will talk more, both about how we are deploying the word Eros, and about existential risk, somewhat in the Corona Afterword and in more depth in the major essays in this volume, particularly Essays One: Introduction and Two.

But at the very outset two key points need to be made.

## **Effective Altruism**

Let us turn to an idea that has been termed effective altruism. The core of the idea is sourced in a simple philosophical principle called consequentialism. This principle suggests that the guide for moral action should not be how it makes you feel but the actual positive effect, the direct consequences of your action. Like most philosophical principles, it is true but partial. It fails to distinguish between in-depth clarified feelings and surface feelings. It treats feelings as mere biochemical events and part of a deeper field of value. That and other important objections to consequentialism notwithstanding, it is a clarifying prism. Effective altruism, in its wake, argues that if the goal of altruism is for example to save lives, then funds should be given in a way that saves most lives. Consequentialism and effective altruism have thus argued that giving money locally in Western democracies is often a moral mistake because your dollars or pounds etc. are less effective. Those same funds might be spent, for example, far more effectively buying malaria nets in third world countries. From the perspective of saving lives, hence effective altruism, that is the far more correct

and therefore noble ethical choice. There is something clean, compelling, and important in the heart logic of effective altruism.

In its wake, we are, however, going to suggest a next step, indeed a momentous leap in the conversation around what is actually effective in responding to both catastrophic and existential risk.

There is a much more effective mode of effective altruism than even that suggested by its proponents—one that is even more consequential. This next step suggests that we must invest massively in that which is actually paradoxically missing from the effective altruism conversation. This is an investment in the development of Simple First Principles. We will discuss what we mean by Simple First Principles in more depth below but for now—by way of introduction—we trust the following will suffice: Complexity points out that highly complex systems are actually not random in the way the term is usually understood. Rather, coherent complex systems are governed by Simple First Principles that are iterated again and again throughout the system. By contrast, systems that become not complex but complicated and incoherent, generating disease and suffering for most of their members, are based on either the absence of or the distortion of Simple First Principles. The existential risk that we face today is directly related to this failure of frameworks—the collapse of Simple First Principles that are aligned with what philosopher Alfred North Whitehead might have called the inherent value propositions of reality.

But we cannot go backwards. We cannot resort to some imagined bygone time. Rather, we must transcend and include the best of our past and present and articulate a new set of Simple First Principle, which are the ground of a New Story and that become the matrix for a global ethos for our global civilization.

With these simple ideas in our hearts and minds, let us return to our not so hypothetical scenario, originally proposed by Oxford philosopher Derek Parfit. What is worse: the 7.7 billion dead in the devolution from 1 to 2 or the remaining 78 million deaths in the move from 2 to 3? We concluded—counterintuitively—that the latter is far more devastating because it constitutes the erasure of the entire future. Paradoxically, effective altruism in the sense of malaria nets, will help us avoid some of the dimensions of catastrophic risk, saving so very many lives—a self-evidently necessary and insanely noble obligation. However, it seems self-evident that it is only the evolution and articulation of a new set of Simple First Principles that will allow us to save the future. That is the purpose of this book and the Great Library of Simple First Principles of which it is a part.

31.08.2020