Biases for Evil and Moral Perfection

Hans Van Eyghen

Article

Abstract: I argue that deeply ingrained dispositions to do evil do not render moral perfection impossible. I discuss various accounts of moral perfection and the evidence from cognitive (neuro)science that points towards a strong disposition for evil. Afterwards, I discuss three strategies that can allow humans to overcome their evil dispositions. These are: cognitive enhancement, avoiding triggering situations and structural solutions.

Keywords: cognitive bias; evil; moral perfection

1. Introduction

The need for or impossibility of moral perfection is a recurrent theme in philosophy. Most claims in favor of or against the possibility of moral perfection rely on a priori arguments or anecdotal evidence. In this paper, I investigate how empirical evidence on biases towards morally bad behavior drawn from evolutionary biology and cognitive science weighs in on the debate. I argue that while evolved human dispositions for morally bad behavior render moral perfection more difficult, humans can achieve moral perfection through various routes. I conclude that moral perfection remains possible for human beings in spite of deeply ingrained dispositions for evil.

Like other authors discussing the possibility of moral perfection I have in mind a concept of “possibility” closer to metaphysical possibility than to logical possibility. Since the concept of moral perfection does not appear to be internally inconsistent, logical possibility is easily granted. Deeply ingrained biases for evil appear to challenge whether moral perfection is possible for humans (homo sapiens) in this world. I argue that this remains the case.

This paper is structured as follows: In Section 2, I discuss what moral perfection is and what varieties can be distinguished; in Section 3, I discuss the evidence from evolutionary biology and cognitive science for a disposition to do evil in man; in Section 4, I discuss how the evidence challenges the possibility of moral perfection; in Section 5, I discuss how the challenges can be met and lay out three ways how humans can overcome their dispositions for evil.

2. What Is Moral Perfection?

The amount of discussion over moral perfection is fairly limited in contemporary philosophy. There is considerable debate over related terms like (the possibility of) God’s omnipotence or moral virtue. Because both terms refer to different phenomena or different beings, the discussion is not straightforwardly applicable to moral perfection. In this section I make an attempt at defining “moral perfection”. In later sections I assess whether humans can ever achieve moral perfection in light of recent advances in evolutionary biology and cognitive science.

While the need for or importance of moral perfection is rarely discussed by contemporary ethicists and other philosophers, the idea of moral perfection or similar traits is prominent in a number of (religious) traditions. Christian traditions often refer to Matthew 5:48 “Be perfect, therefore, as your heavenly Father is perfect.” to regard moral perfection...
as a personal duty. The catechism of the Roman Catholic Church affirms that all Christians should strive towards the fullness of Christian life and the perfection of charity. Humans are able to do so with help from divine grace (Catholic Church 1994). For Eastern Orthodox churches, moral perfection is part of the doctrine of theosis, a process whereby humans are granted likeness to God or union with God. Wesleyan, Methodist doctrine affirms that sanctification through the work of the Holy Spirit is possible. Sanctification is a gradual, progressive process where humans grow in grace and obedience to God culminating in a state that nears moral perfection. Some non-Christian traditions also accept some form of moral perfection. A large number of Buddhists affirm that humans can achieve liberation from the cycle of rebirth through the eightfold path of right practices (right view, right resolve, right speech, right conduct, right livelihood, right effort, right mindfulness and right Samadhi). Following the right path is often regarded as having the power to transform the individual morally beyond its normal limitations. Stoic ethics allow for the possibility of morally perfect acts performed in the right way with an absolutely rational, consistent and formally perfect disposition (cf. Stephens 2004).

I will be using the term “moral perfection” as signifying a (potential) trait of humans. What I claim therefore has no ramifications for the discussion over God’s omnibenevolence or the moral status of non-human animals. I do so because the question whether moral perfection is attainable is usually applied to humans, and humans appear to be the most interesting moral actors in our world.

Moral perfection is a character trait that humans can or cannot possess. In that sense, moral perfection is similar to moral virtue. Unlike moral virtue, moral perfection is not something humans can possess to a greater or lesser extent. Whereas humans can be more or less morally virtuous, they cannot be more or less morally perfect. As with all perfections, moral perfection consists of possessing something to the greatest possible extent.

Earl Conee distinguishes four forms of moral perfection:

1. Doing everything that is morally right and nothing that is morally wrong.
2. Doing what is supererogatory at every moment.
3. Doing what is morally right without any liability.
4. Doing everything that is morally good with the right frame of mind (Conee 1994).

Conee argues that true moral perfection consists of all four traits. He argues that subjects that merely possess (1), (2), (3), or (4) fall short of moral perfection because there remains a possibility for improvement or being even more moral (Conee 1994).

Susan Wolf defines a moral saint as “a person whose every action is as morally good as possible, a person ( . . . ) who is as morally worthy as can be” (Wolf 1982).

Her definition of moral saint does not state that moral saints never do anything morally wrong. Whether someone achieved moral saintliness depends on how morally worthy one can be. Her account of moral saintliness (which she explicitly equates with moral perfection) therefore allows that a moral saint has to do morally bad things for the greater good or has to make moral trade-offs. Wolf thereby suggests a fifth form of moral perfection:

5. Doing as many morally good acts and as few morally bad acts as possible.

The fifth form of moral perfection raises some questions of vagueness. It is not obvious what lies within the range of possibilities for humans. On certain strands of Calvinism, humans are regarded as totally depraved and incapable of doing what is morally right without divine help. Below, I discuss evidence that in-group favoritism and out-group hostility are very hard (if not impossible) to overcome. On both views, the moral possibilities of men are severely limited. Nonetheless, people could meet Wolf’s conditions for moral sainthood because they are doing as many morally good acts as possible within their possibilities. Wolf, however, adds that the lives of moral saints are dominated by a commitment to improve the welfare of others or of society as a whole (Wolf 1982). Wolf, therefore, appears to have a more positive view of what humans can achieve morally in mind.
Wolf’s account puts far fewer limitations on moral perfection than Conee’s. Because she ties moral saintliness to possibilities, humans can perform a considerable number of evil acts and still be regarded as morally perfect. Only performing evil acts that a subject could have refrained from are causes to not grant them moral perfection. Therefore, only acts that stem from malicious intent or weakness of will preclude moral perfection. On Wolf’s account, people suffering from psychopathy or sociopathy can in some cases be ranked among the morally perfect because their psychological conditions make it impossible for them not to do certain evil acts. Not only does Wolf’s account run against common sense, it also does not call for moral transformation or effort to overcome moral limitations.

I noted that Conee argues that moral perfection must involve all possible moral perfections and therefore must include 1–4. Contrary to Conee, I will mostly be investigating the possibility of (1) in the rest of this paper. I do so because discussion on related topics often relies on this understanding of moral perfection. Since the other forms pose stricter constraints on moral perfection, a negative answer to the question whether moral perfection (1) is possible will imply a negative answer for other forms of moral perfection as well. A positive answer does not necessarily have implications for other forms of moral perfection.

3. Empirical Challenges to Moral Perfection

Some authors defend arguments against moral perfection drawn from intuitions or a priori reasoning. The argument I discuss draws on empirical evidence on the psychological nature of humans. While most research focuses on altruism and (evolutionary) explanations thereof, a growing number of psychologists and cognitive scientists argue that humans have deeply ingrained biases or propensities. Some of these propensities would make humans prone towards doing morally bad acts. While this claim is compatible with the possibility of moral perfection, some authors continue that some propensities to commit evil cannot be overcome because they are deeply rooted within human psychology. Any attempt to achieve moral perfection would therefore be doomed to fail. I return to this last point in the next section.

While there is considerable anecdotal evidence for dispositions for evil, the theories I discuss below provide explanations for why such dispositions came about. They do so by pointing to the adaptive use of having certain biases. These adaptive biases give rise to dispositions for evil as a by-product, or the dispositions for evil have adaptive value in themselves. A common objection to these and other evolutionary explanations is that they lack direct empirical confirmation. Defenders point to how their evolutionary explanation can plausibly account for observed phenomena, like regularities in human behavior, or can predict human behavior. Direct evidence for how natural selection ran its course in selecting human cognitive functions is, however, nearly impossible. The main reason is that current science lacks a clear understanding of how cognitive biases or dispositions are encoded in the human genome. Furthermore, DNA strands of ancient human fossils are usually too degraded to give a clear image of their genome. For these and other reasons, direct evidence is probably too much to ask for.

I will note some criticisms but will make little attempts to critically evaluate the scientific evidence I discuss in this section in great detail. Such an evaluation falls beyond the scope of this paper and outside of my expertise. I note that all approaches I discuss have a lot of traction within their respective scientific communities and are widely discussed and accepted. When evaluating the implications of the evidence on moral perfection, I will proceed as if the conclusions drawn from the evidence are true. I also do not discuss evidence for propensities that are conducive to morally good behavior. While a strong case can be made that natural selection also endowed humans with dispositions for morally good behavior, this need not affect my overall point. Since I aim to assess whether certain dispositions preclude moral perfection, these dispositions need not occur very often. It suffices if humans are on some occasions disposed towards morally bad behavior to raise serious doubts about the possibility for moral perfection. If dispositions for evil turn out to be widespread and affect humans on many occasions their existence might also raise
worries for the possibility of moral virtue building or moral progress. These questions, however, lie beyond the scope of this paper.

Before I discuss some of the scientific evidence in more detail, I want to clarify one more point. The authors I discuss below aim to describe general features of the human mind. They offer evidence in support of biases or propensities that are shared by all or most normally developed and normally functioning humans. They do not argue that some subset of the human population (like psychopaths) suffers from deeply ingrained propensities to do evil, but instead argue that most normally functioning humans develop these propensities throughout their lifetime. The empirical evidence could therefore have ramifications for the general possibility of moral perfection for the vast majority of human beings.

3.1. Evolved Bias for Belief in Moralizing Gods

The first evidence for a deeply rooted propensity for evil states that humans have a natural inclination for a particular form of religious belief, which in turn makes humans prone for evil. A number of authors argue that belief in moralizing gods served an adaptive benefit. Believing that there is a God who monitors human moral behavior and who rewards or punishes humans accordingly would lead to better cooperation. People who believe that moralizing gods are watching their every move would be less likely to free ride and more eager to do their fair share in cooperation. Since cooperation is highly important for the human species, belief in moralizing gods would have been selected for by natural or cultural selection.

The idea of gods as a social monitor is rather congenial to moral perfection. If belief in moralizing gods leads to better cooperation, it helps humans do what is morally right. Some authors noted, however, that there is a flipside. Adherence to moralizing gods leads to organization in religious communities. In the past, the scope of religious communities coincided with scope of cooperation. Nowadays, various religious communities exist alongside one another. Religious communities involve membership. Often members are distinguished from non-members by outward markers (usually bodily markers, symbols or clothing). Membership easily leads to exclusion. As John Teehan notes, outsiders are not invested in the group and have less motivation to cooperate or reciprocate cooperation. They are therefore perceived as free riders more easily and as a danger to the community (Teehan 2010).

Others argue that belief in moralizing gods leads to discrimination of subjects that do not share this belief. If belief in moralizing gods is linked to trustworthiness and refraining from free riding, people will tend to distrust those who do not believe in moralizing gods. Nowadays, most non-believers in moralizing gods are atheists. Therefore, the disposition to believe in moralizing gods could explain discrimination of atheists throughout the world (Gervais 2013).

From its onset, the theory has been subject to a number of criticisms. For example, Inti Brazil and Miguel Farias argue that belief in moralizing gods is better explained by a human motivation to reduce uncertainty than by their role as social monitor (Brazil and Farias 2016). Pascal Boyer and Nicolas Baumard argue that belief in moralizing gods is better explained by increased affluence (Boyer and Baumard 2016). Assessing the plausibility of the theory lies beyond the scope of this paper. I do note that both criticisms do not call into doubt that belief in moralizing gods is widespread and was historically important. Disagreements lie in the reasons why (cultural) evolution selected for belief in moralizing gods. Authors arguing for a connection between belief in moralizing gods and in-group preference draw support from the close connection between such belief and cooperation. If belief in moralizing gods was selected for other reasons, the connection to in-group bias is probably less strong.

Excluding outsiders from cooperation or distrusting outsiders is generally considered morally evil. Excluding outsiders can easily lead to xenophobia or parochial behavior, which is generally considered morally bad behavior as well. If an evolved inclination for religious belief leads to an inclination to exclude outsiders, humans therefore have
an inclination for evil. However, the present situation shows that the inclination to believe in moralizing gods is probably not very strong. A considerable number of people in contemporary societies do not hold beliefs in moralizing gods. Nonetheless, some evolutionary biologists hold that even they suffer from evolved inclinations that foster out-group hostility. I turn to an example next.

3.2. Evolved Limits to Altruism

Robert Dunbar argues that there is an upper limit to stable human group size and associated altruism. Maximal stable group size is determined by the volume of the neocortex. The volume of the neocortex sets the cognitive constraints that determine the maximum number of individuals any animal can keep track of. Group cohesion is in turn maintained by various practices. For nonhuman primates, group cohesion is maintained through social grooming. The amount of social grooming is literally related to group size. The relationship between group size and time spent grooming is a consequence of the intensity of grooming with a small number of key “friendships”. These key friendships function as coalitions that have the purpose of buffering members against harassment by other members of the group. The larger the group, the more harassment and stress an individual faces. The mean group size of a species is related to the mean size of the smaller coalitions. Therefore, the total group size is limited by the number of individuals others can keep track off within the range of their cognitive abilities (Dunbar 1993).

Like nonhuman primates, humans are limited by their cognitive abilities as well. Humans have a significantly larger neocortex (roughly 30% more volume than nonhuman primates) and therefore can keep track of more individuals. Mean human neocortical size predicts that human group sizes range from 107.6 to 189.1. Dunbar notes from surveying a number of hunter-gatherer groups that a lot of tribes have a membership of well beyond 190, often ranging in the thousands. However, tribes often consist of intermediate-level groups with an average number of 148.4. These intermediate-level groups appear to be characterized by interaction on a sufficiently regular basis so that they have strong bonds based on direct personal knowledge (Dunbar 1993).

While the discussion up to this point focused on (the limits of) group sizes in traditional cultures, Dunbar’s surveys evidence that 150 may also be a functional limit on interacting groups in contemporary western industrial societies. Research suggests a negative effect of group size on group cohesion and job satisfaction. Some businesses informally hold on to a limit of 150 for effective coordination of tasks and information flow through direct person-to-person links. Larger companies usually have substructures of smaller size. Studies on the number of different acquaintances in modern urban societies also tend to average on a number around or below 150. Dunbar also notes that most organized professional armies also consist of basic units of about 150 men (Dunbar 1993).

Humans do not maintain group cohesion through grooming like apes do. Grooming roughly 150 people would take up too much time and effort. According to Dunbar’s calculations, it would take up 42% of our time. Instead, humans maintain relationships by other forms of social bonding like ritual behavior or language. Dunbar suggests that language evolved as a “cheap” form of grooming. Language does not only serve to exchange information about one’s environment. It also allows individuals to spend time with their preferred social partners and allows them to acquire knowledge about the behavioral characteristics of other group members (Dunbar 1993).

The limit on core friendship groups imposed by limits in human cognitive abilities in turn lead to limits in the size of societies. Dunbar does suggest that the size of societies could increase because of social dispersal. In dispersed societies, individuals meet less face-to-face and are less familiar with each other. Language also raises the possibility of increasing group size by categorizing individuals into types. Individuals can be identified as belonging to a particular class in virtues of having a particular cue (some piece of clothing or bodily marker). By virtue of this, humans need only learn how to behave towards a general type of individual rather than learning how to behave towards every
individual. This could allow for larger human societies. Dunbar does note that super large states have not proven to be very stable through time. Most large empires eventually collapsed or were maintained by means of oppression. He also notes that larger groups appear to be less cohesive than groups that are smaller than the critical limit (Dunbar 1993).

Dunbar’s evidence has similar implications like the theory on moralizing gods. The evidence suggests that there is an upper limit to human cooperation. While humans can cooperate well enough with people from the in-group of 150, the evidence suggests that most humans will frown upon intense collaboration with non-members. The evidence from nonhuman primates even suggests that interactions with an out-group easily involve harassment and stress.

A recent attempt at replication called the limit of 150 members of stable groups into doubt (Lindenfors et al. 2021). Other critiques pointed to the importance of the quantity of nutrients human brains consume compared to apes (Pfeifer and Bongard 2006), and differences in diet (DeCasien et al. 2017). Dunbar does not take both into account. However, all criticisms call for a revision of the maximal group size for stable altruism. Unless the maximum is revised dramatically, revised numbers would still point to serious limitations to human altruism.

A strong preference for members of an in-group is hard to reconcile with moral perfection. While a limit of human group sizes to 150 would allow for moral perfection in small isolated communities, most human societies are far larger. A morally perfect person should be expected to treat everyone equally well. While preferences for closely affiliated fellow humans might not impede moral perfection (everybody has at least some preferred fellow humans), the ease with which in-group fidelity leads to out-group hostility poses a real problem. I turn to this issue next.

3.3. Evolved Tendency for Group Conflict

The dispositions discussed so far can easily lead humans towards doing morally bad deeds. The implications are, however, still rather limited. Both dispositions imply that humans tend to have preferences for members of the in-group in their choice of cooperating partners and will tend to behave more altruistically towards in-group members. Pascal Boyer argues that propensities for in-group behavior can easily have much more profound implications. I turn to his account now.

Boyer argues that no human population is immune from potential ethnic rivalry and conflict. These can escalate into full-blown civil war and genocide surprisingly easily. This disposition can be explained by the evolutionary roots of human behavior. Ethnic rivalry, or other forms of out-group hostility, are the flipside of the human propensity for cooperation. Humans rely more on cooperation than any other species. Cooperation enables human to divide tasks like gathering food or taking care of offspring within their communities. As a result, humans can be more successful in their endeavors. Because cooperation is so important, a large number of evolutionary biologists and evolutionary psychologists argue that natural selection selected for deeply ingrained, strong intuitions and dispositions for cooperation. These intuitions and dispositions make humans prone to find partners for cooperation (Boyer 2018).

Having strong dispositions to cooperate is rather congenial to moral perfection. Nonetheless, they can easily lead to out-group hostility. They do because relying on cooperation also makes human vulnerable. When tasks are divided among a community, the danger of free riders lurks. Free riders reap the benefits of cooperation but do not contribute anything or little themselves. A high prevalence of free riders undermines trust and lowers the benefits of cooperation for non-free riders. For this reason, humans were also endowed by natural selection with dispositions to be on guard against potential free riders. People monitor commitment and defection to avoid investing resources and efforts in coalitions with free riders. As a result, people are easily mistrusting of newcomers and of people from outsiders their in-groups because they have not proven to be trustworthy
collaborators. Humans often rely on tokens like language, skin color or traditions as cues or markers of belonging to the in-group (Boyer 2018).

According to Boyer, a disposition for mistrusting members of an out-group can easily lead to out-group hostility. To safeguard their own in-group and its associated cooperation, humans easily perceive any interference from out-group members as attacks. Small infringements by individual members of an out-group, like petty thefts or insults, can easily trigger strong defensive reactions. Boyer notes that inter-group conflict often takes a predictable form. In many cases, conflicts start with a minor episode, like a scuffle between youths or an angry reaction to a sporting event. In some cases the minor event is amplified by rumors of deliberate acts of aggression by the out-group. After the rumors have disseminated (and in some cases were amplified by media reports), members of one group grow to be cautious in their interactions with members of other groups. When a new minor event occurs after a few days, the conflict escalates and more serious reactions like storming out-group districts or even killing members of out-groups occur (Boyer 2018).

Like other evolutionary psychologists, Boyer argues that human cooperation and the human dispositions for in-group preference and out-group hostility are the result of unconscious computations. While the reason for the strong dispositions is often not accessible through introspection, they have a strong effect on human behavior when triggered in the right way. The dispositions were selected for by natural selection because cooperation is very important and because free riding poses a serious threat for human survival (Boyer 2018).

The effects of an inclination for out-group hostility are clearly morally bad. Whereas the moral effects of limits to altruism were still rather limited to xenophobia or distrust of strangers, the effects of the dispositions discussed by Boyer are much more far-reaching. It is clear that a disposition for aggression towards members of the out-group is morally bad behavior.

4. The Impossibility of Moral Perfection

While some authors defended a priori reasons for doubting the possibility of moral perfection (e.g., Conee 1994), the evidence I discussed in the last section gives empirical reasons to doubt that humans can achieve moral perfection. The evidence strongly suggests that humans have strong dispositions for morally bad behavior like preferences for in-group altruism and out-group hostility. The evidence also suggests that the dispositions are triggered easily so that a lot of humans will suffer from their effects.

Having strong dispositions for morally bad behavior does not render moral perfection impossible but merely renders it difficult. Through virtue development, exercise of the will or other means, humans might be able to overcome their natural dispositions and learn to avoid morally bad behavior. Overcoming evolved dispositions will, however, prove to be difficult for a number of reasons. First, as Boyer noted, some dispositions consist of unconscious mental computations. Because these are not open to introspection, most humans are not aware of the dispositions and triggering conditions. Lacking information about dispositions for morally bad behavior and how they work will prevent humans from finding the means to overcome the dispositions. This problem can be solved by making information more widely available. By instructing people about their dispositions for morally bad behavior, people can learn how their dispositions are triggered and take steps towards overcoming them.

A second problem runs deeper. There is ample evidence that dispositions rooted in evolved psychology tend to resurface when subjects lack the time or resources for adequate reflection. Some authors suggested that young children are prone to answer why-questions teleologically (e.g., Kelemen 1999). The bias recedes when humans learn mechanistic, scientific explanations for facts. However, when put under time pressure, even trained scientists are again prone to give teleological answers (Casler and Kelemen 2008). Another evolved disposition is fear of snakes (cf. Öhman and Mineka 2003). Information about snakes or reflection can remove fear of snakes in safe conditions. Nonetheless, humans are
still prone to have fear reactions when they encounter a snake or something resembling a snake off-guard. A final evolved disposition is our craving for sugar and fatty food. Eating as much sugar and fatty food as possible made evolutionary sense when calories were scarce. Nowadays, humans can overcome these cravings and choose to eat more healthy food. On occasion (e.g., under the influence of stress or anxiety), many do slip back into preferring sugar or fatty food.

The evidence suggests that although humans can overcome evolved dispositions on many occasions, the dispositions tend to resurface when humans let down their cognitive guard and act unreflectively. Given that it is near impossible to never let one’s guard down, we should not expect that the dispositions for morally bad behavior will never resurface. The question remains how often the dispositions will resurface when humans put in the effort to overcome or suppress them. Considering that we are investigating the possibility of moral perfection, they need not resurface often.

5. The Possibility of Moral Perfection

Having discussed evidence for a strong tendency towards out-group hostility, I will now discuss various pathways how moral perfection might still be possible in spite of deeply ingrained propensities to do evil. I will not challenge the empirical evidence I discussed in Section 2 and proceed as if humans have a deeply rooted propensity for out-group hostility that easily resurfaces when humans let down their cognitive guard. I discuss the strengths and weaknesses of three strategies.

5.1. Cognitive Enhancement

Dunbar’s evidence for a correlation between maximum group size and cognitive abilities raises the suggestion that maximum group size can be increased by having larger cognitive abilities. Being able to compute a larger group size might also help humans in expanding the in-group and limiting out-group hostility. Usually cognitive abilities in species change as a result of mutations in the genome, which are transmitted if they lead to increased fitness. Waiting for natural selection to select for increased cognitive abilities would, however, take too much time and would not help humans achieve moral perfection in the short run. Humans might intervene in the evolutionary process by selecting individuals with higher cognitive abilities and having them reproduce. Such practices are, however, not deemed morally acceptable and are condemned in most contemporary codes of law.

Cognitive enhancement, however, need not imply intervention in the human genome. Some authors suggest that cognitive enhancement is possible by pharmacological or other means. These are able to alter not just human action but also human dispositions. As Bostrom and Sandberg note, cognitive intervention can be both therapeutic (where a pathology or impairment is cured) or enhancing (where human abilities are increased) (Bostrom and Sandberg 2009). Given the evidence cited in Section 2, it appears that humans need enhancing intervention to obtain increased cognitive abilities. One form of cognitive enhancement is psychological interventions like learned tricks or mental strategies. Basic examples are education and training with the goal of improving mental faculties like concentration or memory. Other forms of psychological interventions move beyond these, like mindfulness training, yoga, meditation or martial arts (Bostrom and Sandberg 2009).

Other forms of cognitive enhancement involve drugs or other substances. Bostrom and Sandberg note that humans have been using caffeine to increase attention for centuries (Bostrom and Sandberg 2009). Some have suggested that medicines used for the therapeutic treatment of Attention Deficit Hyperactivity Disorder (ADHD), like methylphenidate, dexamphetamine, lisdexamfetamine, atomoxetine and guanfacine, could also increase human capacities for attention in non-therapeutic contexts.

A worry for using drugs to increase human cognitive abilities is that not much is known about their potential (side) effects and their long-term consequences (Bostrom and Sandberg 2009). Drugs that have a lasting effect on the brain might also raise issues
about bodily integrity and even personal identity. More importantly for our purposes, it is unclear whether drugs can solve the limitations in neocortical volume, which leads to limitations in group size according to Dunbar. While drugs can alter the operations of neurotransmitters or might impact brain activity, it is doubtful that they can increase the volume of the human neocortex.

Another form of cognitive enhancement could be more promising. Bostrom and Samberg note that cognitive enhancement can also be achieved by external hardware and software systems. Humans have been using calculators or personal computers to increase their cognitive abilities for quite some time. Software can help displaying information, keep more items in memory and perform routine tasks (like mathematical calculations) more rapidly. Data mining and information visualization make large quantities of data easier to handle (Bostrom and Sandberg 2009).

Cognitive enhancement by means of hardware or software can come in two forms. In its classical form, humans use an external device (e.g., a calculator, PC or personal digital assistant) to help them perform cognitive tasks. At any point in time, the user can decide to turn the device on or off. Despite being external to the human brain, devices can have a tremendous effect on human cognitive abilities. More recent, internal hardware enhancements became possible. Humans can have electrodes implanted in their brain. The hardware can make use of software that can interpret incoming sensory signals and commands (Bostrom and Sandberg 2009).

If humans are to increase their cognitive abilities to allow for (far) greater group size, much more is needed than hardware and software that merely imitates human cognitive behavior. Hardware and software must be developed that surpasses human abilities, and the hardware and software must be coupled to human brains in an external or internal way. Current development in artificial intelligence still seems far removed from this goal. However, there is no a priori reason to think that the necessary advances cannot be achieved. In any case, increasing cognitive abilities by means of hardware and software is still science fiction and not of any help for humans to achieve moral perfection in the near future.

5.2. Avoiding Triggers

Another solution does not put its hope on technological developments but on practices humans can adopt to avoid the nefarious effects of dispositions for evil. Knowing about their cognitive limitations and the effects they can have on treatment of outsiders, humans can choose to avoid contact with individuals beyond their core-group of roughly 150. By doing so they avoid the stress and potential harassment than easily comes with contacting outsiders. In a similar way, they could avoid situations that trigger dispositions for out-group hostility.

Consider this example:

Rupert has an innate propensity to steal jewelry. On the road from Rupert’s home to work is a jewelry store. Whenever he walks past the store, Rupert is triggered to go in and try to steal some of the jewelry on display. Though having tried to contain his urges, Rupert finds himself unable to refrain from trying to steal the jewelry. To avoid actually stealing any more jewelry, Rupert decides to take a different route to work. He no longer passes by the jewelry store every day and tries to avoid passing by it on other occasions as much as he can.

Rupert’s solution to avoid doing evil is somewhat similar to the psychological interventions we discussed above. Rather than trying to change his propensities (what psychological interventions often try to do), Rupert changes his behavior in such a way that he is no longer triggered to do evil acts. His dispositions are still there. If Rupert would decide to take the old route to work, he would again be triggered to steal. By changing his behavior to avoid the trigger, he manages to avoid the real-world effects of his propensities.

In a similar way, humans could avoid the effects of their disposition towards out-group hostility by avoiding the trigger. If humans are able to keep their contact largely within their core-group, they could avoid the effects of their limited cognitive abilities. They can
happily cooperate with people within their core-group and not be bothered by outsiders. Because of the lack of contact with outsiders, they are not triggered to treat outsiders in a bad way and are therefore not triggered to do morally bad acts.

The strategy seems to suffice to achieve moral perfection (1). By not being triggered, humans refrain from doing anything bad and focus on the good they can do within their core-group. They, however, fail to achieve moral perfections 2–4.

A worry for this solution is whether limiting contacts to the in-group is always possible. People living in large urban areas or practicing jobs that involve meeting a lot of people will be unable to limit their engagements to their in-group. Increased interconnectivity by means of the internet and social media also increase the number of contacts (far) beyond 150. The contemporary, globalized society therefore calls into doubt whether the second strategy is tenable.

5.3. Structural Solutions

A last solution that can enable moral perfection despite deeply rooted propensities for evil is building top-down structures to avoid the consequences of propensities. Like in the previous solution, subjects are prevented from doing evil by avoiding triggers. Unlike the previous solution, the subject does not avoid triggers in a bottom-up way but is prevented from encountering triggers by structures put in place in a top-down way by the state, organizations or others.

Structural solutions are widely discussed as a solution for (evolved) biases like those that lead to collective action problems (situations where society would be better off by cooperating but individuals fail to do so because of conflicting interests between individuals that discourage joint action). Some proposed introducing procedures that nudge or even force subjects towards cooperation. Subjects can be nudged by offering monetary or other rewards or by presenting more information about the effects of cooperation.16

Another domain where structural solutions have been proposed to help subjects overcome biases in judgment is medicine. Schleger, Oehninger and Reiter-Theil suggest using checklist as a mean to help subjects become aware of biases in medical decision-making and to adjust their judgments accordingly (Schleger et al. 2011). Other means of overcoming biases in the medical domain are crosschecking of diagnoses by multiple doctors. Similar protocols have also been put in place in other domains of medicine.

Protocols that automatically kick in when needed could be put in place to avoid a distrusting reaction towards outsiders or to avoid violent reactions. Considering that distrusting reactions towards outsiders appear to be triggered very easily, avoiding violent reactions seems more manageable.

What could structural solutions that avoid triggers for the dispositions I discussed in Section 3 be like? Dunbar gives us a cue to a solution for the cognitive limitations for altruism. He argues that multiple individuals can be represented as a type. By using types, human minds can keep track of more people. For example, by categorizing individual police officers as “police agent”, humans do not need to store individual police officers in their mind but can rely on a type. As a result, humans behave similarly towards police officers (Dunbar 1993). By having people classify more people as various types, education can help humans keep track of more individuals in their mind and increase the scope of altruism.

Another structural solution to prevent out-group hostility could consist of institutions for inter-cultural or inter-group dialogue. Historical examples of institutions to guide inter-group dialogue are the South African Truth and Reconciliation Commission and Rwandan National Unity and Reconciliation Commission. Both commission had rather different goals, namely reconciling two or more groups after years of conflict. The institutions I have in mind would be preventing inter-group conflict rather than healing or reconciling groups. A preventive institution would need to focus less on forgiveness or healing past trauma and more on discussing and resolving grievances before they escalate. The closest thing in existence that resembles such an institution are the parliaments of Lebanon, India and New
Zealand. In Lebanon, seats are allocated to various religious groups to prevent dominance of one group. In India, a number of seats are allocated to members of the Dalits. The New Zealand parliament reserves a number of seats for members of Maori groups. Since parliaments usually have rather different goals than discussing grievances (i.e., legislating), other kinds of institutions seem preferable. Furthermore given that inter-group conflict usually escalates at a local level, more localized institutions seem preferable.

Discussing how such institutions should be construed in detail lies beyond the scope of this paper. I end with some general reflections on how the institutions can proceed. Because of the high emotional salience of inter-group behavior (as laid bare by Boyer), conversation should be structured in a non-violent, deescalating way. People participating in conversations should also be mindful of the history of interactions and conflicts between all participating groups. Lastly, participants should be aware of the various biases and cognitive limitations that foster inter-group hostility.

6. Conclusions

I argued that evidence for deeply ingrained biases for evil that are hard to overcome do not give us sufficient reason to doubt that moral perfection is impossible. I discussed evidence pointing towards deeply ingrained dispositions to do evil in humans. I also pointed to a number of ways how humans can overcome these limitations.

Showing that humans can overcome these propensities, however, does not suffice to show that humans can achieve moral perfection. Other propensities or biases that lead to evil acts could still be at work, and it is not clear whether the strategies I discussed will help humans in overcoming these. There could also be other (a priori) reasons to doubt the possibility of moral perfection. My arguments do suffice to show that a deeply ingrained propensity for evil need not definitively preclude humans from achieving moral perfection.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1. Most of the discussion whether divine omnibenevolence is possible boils down to questions over the logical possibility of moral perfection since God is not bound to a fallible mind or worldly influences like humans are. See for example: (Garcia 2009). Other questions concerning divine omnibenevolence are whether divine omnibenevolence is compatible with a wide prevalence of evil (e.g., Rowe 1979).

2. Achieving moral perfection is not the main goal of Buddhist spiritual practice. The main goal is achieving liberation or enlightenment. Nonetheless, moral transformation appears to be an indispensable part of what Buddhists are striving for.

3. Wolf’s main thesis is that moral saintliness puts too high a burden on people (Wolf 1982). I will not be concerned with this debate throughout this paper.

4. See for example: (Conbee 1994).

5. I also don’t discuss issues concerning chance in human evolution (see for example: Alexander 2020) or criticisms of the core tenets of evolutionary psychology (see for example: Ketelaar and Ellis 2000).

6. See for example: (De Waal 2008; Pinker 2011).

7. The theory comes in two varieties. The first maintains that belief in moralizing gods is a biological adaptation (Johnson 2015). On this theory religious belief was selected for when our ancestors lived in the African savannah and transmitted through human genetic material. The second variant maintains that belief in moralizing gods is a cultural adaptation. On this theory, belief only began to serve an advantage when our ancestors began living in large-scale communities during the Neolithic revolution. Around that time, groups with belief in moralizing gods outcompeted other groups (Norenzayan 2013).

8. Although some of Boyer’s other views have been subject to considerable criticism (e.g., Sterelny 2018; Gregory and Greenway 2017), his more recent views on evolutionary explanations for out-group hostility have not so far.

9. See for example: (Tooby and Cosmides 2010).

10. William Swann and others argue that processes of identification with group members can explain why interferences or attacks on group members are often perceived as personal attacks (Swann et al. 2012).

11. A clear example to which Boyer refers are the Nika Riots in Constantinople 532.
Lari Launonen makes a similar argument stating that cognitive limitations might prevent humans from achieving cognitive perfection (Launonen forthcoming).

Another possibility that I will not discuss is that humans can overcome evil dispositions by some process of transformation induced by God. Many Christians believe that humans will be transformed by God in the eschaton. This transformation could very easily include a transformation of the mind and its dispositions. For example, Alvin Plantinga argues that (some) human cognitive dispositions can be transformed by the intervention of the Holy Spirit (see: Plantinga 2000). Proposals like these are usually wedded to the idea that moral perfection is impossible for humans without divine help. My arguments, if successful, show that a form of moral perfection can be achieved by humans themselves.

Biologists also allow that mutations can spread as a result of genetic drift or as a by-product of other adaptations.

Some also discussed the possibility or duty to use attention-increasing drugs for pilots or surgeons (see: Kloosterboer and Wieland 2017). The goal is increasing attention to perform certain tasks in a better, more responsible way and not for epistemic or moral purposes.

See (Ostrom 1990) for a discussion of procedures of this kind.

See for example: (Rosenberg 2002).

References

Alexander, Denis. 2020. Is Evolution a Chance Process? Scientia et Fides 8: 15–41. [CrossRef]

Bostrom, Nick, and Anders Sandberg. 2009. Cognitive Enhancement: Methods, Ethics, Regulatory Challenges. Science and Engineering Ethics 15: 311–41. [CrossRef] [PubMed]

Boyer, Pascal. 2018. Minds Make Societies. How Cognition Explains the World Humans Create. New Haven: Yale University Press.

Boyer, Pascal, and Nicolas Baumard. 2016. Projecting WEIRD Features on Ancient Religions. Behavioral and Brain Sciences 39: 1–65. [CrossRef] [PubMed]

Brazil, Inti A., and Miguel Farias. 2016. Why Would Anyone Want to Believe in Big Gods? Behavioral and Brain Sciences 39. [CrossRef]

Casler, Krista, and Deborah Kelemen. 2008. Developmental Continuity in Teleo-Functional Explanation: Reasoning About Nature Among Romanian Romani Adults. Journal of Cognition and Development 9: 340–62. [CrossRef]

Catholic Church. 1994. Catechism of the Catholic Church. Vatican City: Libreria Editrice Vaticana.

Conner, Earl. 1994. The Nature and the Impossibility of Moral Perfection. Philosophy and Phenomenological Research 54: 815–25. [CrossRef]

De Waal, Frans B. M. 2008. Putting the Altruism Back into Altruism: The Evolution of Empathy. Annu. Rev. Psychol. 59: 279–300. [CrossRef]

DeCasien, Alex R, Scott A. Williams, and James P. Higham. 2017. Primate Brain Size Is Predicted by Diet but Not Sociality. Nature Ecology & Evolution 1: 1–7.

Dunbar, Robin I. M. 1993. Coevolution of Neocortical Size, Group Size and Language in Humans. Behavioral and Brain Sciences 16: 681–94. [CrossRef]

Garcia, Laura L. 2009. Moral Perfection. In The Oxford Handbook of Philosophical Theology. Edited by Thomas Flint and Michael Cannon Rea. Oxford: Oxford University Press.

Gervais, Will M. 2013. In Godlessness We Distrust: Using Social Psychology to Solve the Puzzle of Anti-Atheist Prejudices. Social and Personality Psychology Compass 7: 366–77. [CrossRef]

Gregory, Justin P., and Tyler S. Greenway. 2017. Is There a Window of Opportunity for Religiosity? Children and Adolescents Preferentially Recall Religious-Type Cultural Representations, but Older Adults Do Not. Religion, Brain & Behavior 7: 98–116.

Johnson, Dominic P. 2015. God Is Watching You: How the Fear of God Makes Us Human. New York: Oxford University Press.

Kelemen, Deborah. 1999. The Scope of Teleological Thinking in Preschool Children. Cognition 70: 241–72. [CrossRef]

Ketelaar, Timothy, and Bruce J. Ellis. 2000. Are Evolutionary Explanations Unfalsifiable? Evolutionary Psychology and the Lakatosian Philosophy of Science. Psychological Inquiry 11: 1–21. [CrossRef]

Kloosterboer, Naomi, and Jan Willem Wieland. 2017. Enhancing Responsibility. Journal of Social Philosophy 48: 421–39. [CrossRef]

Launonen, Lari. Forthcoming. Cognitve Regeneration and the Noetic Effects of Sin: Why Theology and Cognitive Science May Not Be Compatible. European Journal for Philosophy of Religion.

Lindenfors, Patrik, Andreas Wartel, and Johan Lind. 2021. ‘Dunbar’s Number Deconstructed. Biology Letters 17: 20210158. [CrossRef] [PubMed]

Norenzayan, Ara. 2013. Big Gods: How Religion Transformed Cooperation and Conflict. Princeton: Princeton University Press.

Öhman, Arne, and Susan Mineka. 2003. The Malicious Serpent: Snakes as a Prototypical Stimulus for an Evolved Module of Fear. Current Directions in Psychological Science 12: 5–9. [CrossRef]

Ostrom, Elinor. 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge: Cambridge University Press.

Pfeifer, Rolf, and Josh Bongard. 2006. How the Body Shapes the Way We Think: A New View of Intelligence. Cambridge: MIT Press.

Pinker, Steven. 2011. The Better Angels of Our Nature: The Decline of Violence in History and Its Causes. London: Penguin Books.

Plantinga, Alvin. 2000. Warranted Christian Belief. New York: Oxford University Press.

Rosenberg, Marshall B. 2002. Nonviolent Communication: A Language of Compassion. Encinitas: Puddledancer Press Encinitas.

Rowe, William L. 1979. The problem of evil and some varieties of atheism. American Philosophical Quarterly 16: 335–41.
Schleger, Heidi Albisser, Nicole R. Oehninger, and Stella Reiter-Theil. 2011. Avoiding Bias in Medical Ethical Decision-Making. Lessons to Be Learnt from Psychology Research. Medicine, Health Care and Philosophy 14: 155–62. [CrossRef]

Stephens, William O. 2004. Stoic Ethics: Epictetus and Happiness as Freedom. London: Continuum.

Sterelny, Kim. 2018. Religion Re-Explained. Religion, Brain & Behavior 8: 406–25.

Swann, William B., Jr., Jolanda Jetten, Ángel Gómez, Harvey Whitehouse, and Brock Bastian. 2012. When Group Membership Gets Personal: A Theory of Identity Fusion. Psychological Review 119: 441. [CrossRef] [PubMed]

Teehan, John. 2010. In the Name of God: The Evolutionary Origins of Religious Ethics and Violence. Oxford: Wiley Online Library.

Tooby, John, and Leda Cosmides. 2010. Groups in Mind: The Coalitional Roots of War and Morality. In Human Morality and Sociality: Evolutionary and Comparative Perspectives. London: Macmillan International Higher Education, pp. 91–234.

Wolf, Susan. 1982. Moral Saints. The Journal of Philosophy 79: 419–39. [CrossRef]