261 *Etica & Politica / Ethics & Politics*, XXIII, 2021, 2, pp. 261-282
ISSN: 1825-5167

# THE EXECUTION HYPOTHESIS FOR THE EVOLUTION OF A MORALITY OF FAIRNESS

RICHARD WRANGHAM
*Department of Human Evolutionary Biology*
*Harvard University*
wrangham@fas.harvard.edu

ABSTRACT
Humans are both the only species known to have a morality of fairness, and the only species in which the social hierarchy is headed by an alliance (a 'reverse dominance hierarchy'). I present evidence in support of the argument by Boehm (1999, 2012) that these two features are causally linked. The reverse dominance hierarchy is detectable in the fossil record around 300,000 years ago with the origin of *Homo sapiens*. From then onwards, according to the execution hypothesis, an alliance of adult males held the power of life and death over all members of the social group, and they used this power to advance their interests. The result was an intense selective pressure against antisocial behaviour and in favour of prosociality, cooperation and conformity to group norms, whether the norms were beneficial for the group as a whole or merely for the male alliance. The execution hypothesis thus argues that group dynamics have operated for at least 12,000 generations to favour the evolution of moral emotions, many of which are designed to protect individuals from the threat of severe punishment or death at the hands of a dominant alliance of males.

KEYWORDS
Self-domestication; reverse dominance hierarchy; alpha male; patriarchy; self-protection

The question that motivates this paper, following Korsgaard (1996), is why humans tend to behave morally: What logic explains to an observer why an agent feels that s/he has to follow a given moral rule? The answer to be given here addresses the existence and nature of emotions that motivate moral decisions. It comes from an evolutionary scenario (the execution hypothesis) developed by Boehm (1999, 2012, 2018) and Wrangham (2019, 2021). The execution hypothesis purports to explain why the human moral system contains the following features, none of which are known to occur in other species.

1. Group norms that treat behavioural acts as right or wrong.
2. Individuals tending to treat many social behaviours as being right or wrong.
3. Communally approved punishments of norm violators.
4. Individuals tending to conform to norms even when doing so incurs personal costs (such as hurting oneself or one's kin).
5. A tendency for moral norms to favour male interests.

The first four features, in combination, explain why evolved human emotions cause agents to tend to act morally. Briefly, individuals are born into communities in which certain behavioural acts are socially categorized as right or wrong, and in which wrong acts are subject to being punished. Group members grow up learning what is considered to be right or wrong, and are emotionally primed with a tendency to conform to group norms. The evolutionary question is why the four features are found in *Homo sapiens* and not in other species.

I draw attention to the fifth feature for both its theoretical significance and its social-cultural importance. The execution hypothesis claims that the human moral system emerged out of male political dynamics in a way that came to benefit males belonging to a powerful alliance more than it benefited other group members. This evolutionary reconstruction is intriguingly complemented by the observation that a universal feature of human society is the presence of institutions that have large influences on moral norms, especially law and religion, which are organized principally by males and which often favour male interests more than female interests (Hudson et al., 2020). To discuss the evolution of morality without considering why it includes some strongly patriarchal elements would therefore be a critical omission.

Emotions are clearly important in moral judgment, given that agents can be committed to moral decisions for which they are unable to produce any rational explanation (i.e. moral dumb-founding, Haidt, 2012). Furthermore neural regions have been identified that are more engaged in the production of quick and automatic emotional responses than in slower, consciously reasoned reactions (Greene and Young, 2020). Such emotions are theorized to have an innate component in the form of a norm psychology, i.e. a tendency to acquire norms, comply with norms, and punish norm violators (Chudek and Henrich, 2011; Sripada and Stich, 2006). The norms themselves are evidently not innate, however, given that they vary among populations and are acquired by individuals during life.

In this paper I do not address the cognitive processes that integrate an agent's emotional and rational responses, the developmental experiences that shape the agent's sense of self or right and wrong, the way that cultural, social, historical, ecological or other factors influence the evolution of the norms in any specific society, or the

conscious rationalization of moral decisions. Such topics are necessary for a full answer to the normative question, because they explain individual and societal variation in moral tendencies and moral categorization. My goal is limited to hypothesizing why humans as a species are psychologically adapted for creating and engaging in a moral system, and how that moral system evolved.

The version of morality that the execution hypothesis addresses is the morality of fairness (concerned with responsibility, obligation and duty, for example) rather than the morality of sympathy (concerned with compassion, concern and benevolence). While elements of a morality of sympathy occur in non-humans, a morality of fairness is not known in any species other than humans (Andrews, 2020; Burkart et al., 2018; de Waal, 2006; Tomasello, 2016). Perhaps the strongest candidate for a nonhuman morality of fairness has been an apparent tendency for inequity avoidance, suggested when captive capuchins *Cebus apella* or chimpanzees *Pan troglodytes* refuse to eat food items of lower quality than a peer receives (Brosnan and de Waal, 2014). However such sacrifice of personal gain is better understood as an effort to manipulate a human experimenter rather than as a protest about inequity (Engelmann et al., 2017; McAuliffe and Santos, 2018). All subsequent references to morality refer to the morality of fairness.

I begin by summarizing, without supporting evidence, the scenario for how the execution hypothesis portrays the evolution of morality. Next I briefly illustrate the empirical evidence and theoretical inference that generates this scenario. I then comment on the status of the hypothesis in comparison to other evolutionary analyses, and on its implications for understanding why questions of morality are often biased towards male interests. Finally I discuss how the hypothesis contributes to answering the normative question.

## SCENARIO FOR THE ORIGIN AND EVOLUTION OF A MORALITY OF FAIRNESS.

The following scenario comes primarily from Boehm (2012) and Wrangham (2019), as summarized by Wrangham (2021).

### *1. Evolutionary background.*

Half a million years ago the immediate ancestors of *Homo sapiens* were *Homo heidelbergensis*, a species that is inferred to have lived, like all non-*sapiens* species, without moral rules. *H. heidelbergensis* occupied open wooded countryside in many parts of Africa in independent groups that were probably rather small, perhaps

averaging 20-30 individuals. They were hunter-gatherers reliant on fire and tools for their survival, sleeping in temporary campsites. Within groups, as in chimpanzees and gorillas, all adult females were socially subordinate to all adult males; and one male, the alpha, dominated the group by virtue of his having defeated all challengers in one-on-one fights. As occurs in non-human primates the alpha could be compassionate, tolerant and cooperative, especially when his interests were aligned with those of others, such as in interactions involving mating, kinship, or mutualistic acquisition of food. When his interests conflicted with those of others, by contrast, he would typically act selfishly, for example monopolizing resources so far as he could, sexually coercing unwilling females, and responding aggressively to males who competed with him for status or resources. The alpha male thus maintained his dominant status by being a bully, reacting with rapid and powerful aggression to any challenge to his authority.

With regard to other relationships, mothers likely cooperated with each other to some extent in activities such as parenting, foraging, cooking, or the making and using of tools. Males cooperated in the contexts of hunting, intergroup aggression, and limited sharing of resources such as food and tools. Male-female social relationships might or might not have included longterm bonds. Some form of language was present, but it was too crude to allow individuals to cooperate in creative ways by sharing each other's intentions in detail. Relationships between groups were likely mostly tense, but they allowed for adolescent females to transfer between groups.

*2. The emergence of a reverse dominance hierarchy and self-domestication.*

Between 400,000 and 300,000 years ago a transformative social dynamic emerged. By virtue of a more sophisticated version of language than existed previously, subordinate males ("beta males") became able to conspire in such a way that they could safely and deliberately kill their group's alpha. Lethal aggression thus became cheap.

Males with domineering tendencies were killed even if they were close kin to the conspirators. Consistent practice of tyrannicide eliminated the alpha role and created a new style of dominance hierarchy (a "reverse dominance hierarchy") such that the alliance of beta males became the dominant power in the troop. The predictable deaths of those that attempted to behave in the domineering style of alpha males now meant that selection acted against reactive aggression. The resulting increase in docility was accompanied by the evolution of numerous features of anatomy, physiology, and cognition that occur in other domesticated or self-domesticated species, and changed *H. heidelbergensis* into a self-domesticated human, *H. sapiens*. In the subsequent ~12,000 generations from then to the present, face-to-face aggression increasingly fell in frequency and intensity, while within-group cooperation increased. Within the male

alliance that suppressed alpha-male behaviour, the threat of being killed for being too aggressive meant that dominance relationships became strictly egalitarian.

Other social relationships also changed at this time. In the new system every group was part of a language-based network of perhaps ten or more groups. To judge from contemporary patterns, groups had generally peaceful relationships with others speaking the same language, but relationships with people speaking a different language would have been tense or hostile. Sexual partners could be found in other groups, and sexual bonds became equivalent to marriage.

*3. The emergence of a moral system.*

The beta males' ability to eliminate the most individually dangerous member of their group meant that they could equally well kill anyone they chose. This was an evolutionarily novel phenomenon: a nonhuman alpha such as a male chimpanzee bullies other adults but cannot deliberately kill them. The alliance's new power of life and death over all group members created an intense selective pressure in favour of the males' shared interests. Violators of the males' interests were threatened, subordinated and/or killed.

Values that favored the interests of the male alliance could be beneficial for the group as a whole, such as cooperation being "good", and murder (unless sanctioned by themselves) being "bad". Other norms were good for the male alliance but not necessarily for the group as a whole, such as male dominance over females being "good" and insubordination by young males being "bad." Enforcement of such norms led to the evolution of reduced antisocial and non-conformist behaviour, and increased prosociality and cooperation. Moral emotions evolved accordingly. Demonstrative feelings of guilt, shame or embarrassment helped to protect innocents from accusations that they had challenged group norms. Sensitivity to others' perceptions helped promote a good reputation.

In sum, killing of those who did not conform to the interests of the dominant male alliance occurred sufficiently often to create a selection pressure in favour of both self-domestication and a moral psychology (Fig. 1). The logic that "explains to an observer why the agent feels that he has to do this [moral act]" is accordingly that his or her moral emotions have evolved to maintain a good reputation as a conforming member of his or her group. The four key components listed above (group norms, a sense of right and wrong, punishment of wrong-doers, and conformism) were thus generated by a combination of the exertion of coalitionary power by senior males and the self-protective behaviour of all group members.

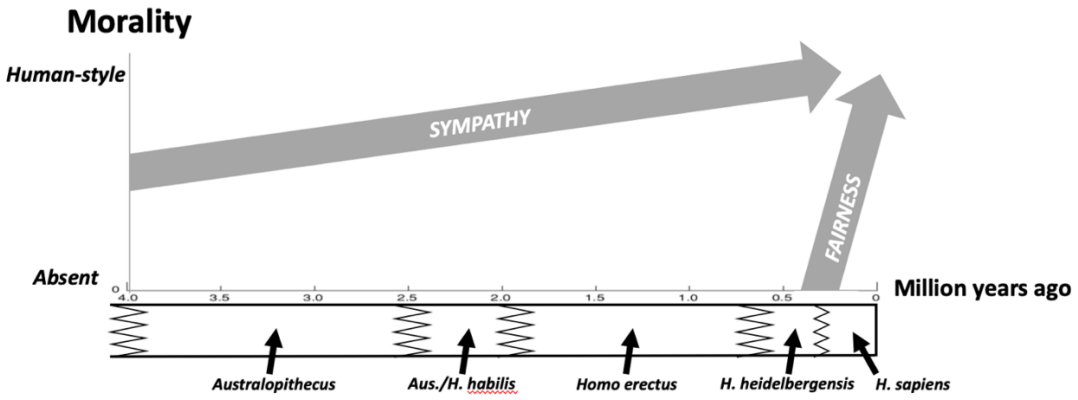


Figure 1. Idealized concept of the evolutionary trajectories of two types of morality. The morality of sympathy is assumed to have a long evolutionary trajectory from pre-human and pre-hominin ancestors. The morality of fairness is proposed to have originated in the transition from *Homo heidelbergensis* to *H. sapiens*.

## EVIDENCE AND INFERENCE FOR THE ABOVE SCENARIO.

The fact that behaviour does not fossilize means that any evolutionary explanation of morality is necessarily speculative. Nevertheless, the execution hypothesis is prompted and supported not only by primatological and ethnographic observations but also by recent anatomical and genetic evidence for human self-domestication. Evidence for self-domestication contributes to the reconstruction of moral origins by identifying the time when the alpha male role was eliminated, i.e. when the evolutionary value of being an alpha male was reversed from positive to negative (Fig. 2).

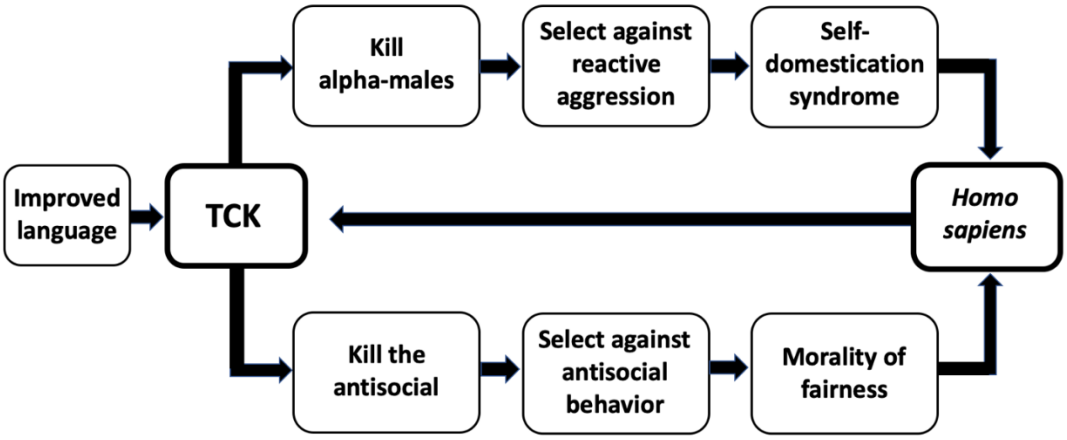


Figure 2. Hypothesized effects of targeted conspiratorial killing (TCK). Following Boehm (2012, 2017) and Wrangham (2019), the top line shows the evolution of self-domestication. Self-domestication is expected to occur in many species, but the purported mechanism that produced it in humans (killing alpha males) is unique to humans. The bottom line shows the evolution of the morality of fairness, which is also unique to humans. Targeted conspiratorial killing had its effects because it enabled kills to be conducted cheaply, i.e. at low risk to the killers. Increases in social tolerance and cooperation made targeted conspiratorial killing increasingly easy to organize. Figure is modified from Wrangham (2021).

Data in support of the above scenario have been reviewed in detail elsewhere (Boehm 1999, 2012, 2017, 2018; Gintis et al., 2015; Hare 2017; Wrangham 2018, 2019, 2021). Here I briefly summarize core components. The numbered paragraphs are merely a list. They do not correspond to the above stages of the evolution of morality.

1. A critical starting-point for understanding moral origins is the question of how pre-moral dominance relationships among adult males transitioned into a moral system. This claim stems from the fact that all organisms have conflicting interests, and in groups of non-human animals such as primates, conflicts are resolved by a dominance hierarchy headed by the best individual fighter. In small-scale societies of *H. sapiens*, by contrast, the dominance hierarchy is unique in being headed by a multi-male alliance within which relationships among males are egalitarian (Flanagan 1989). Conflicts of interest are resolved by a system of moral rules that reward upholders of the moral system and punish those committing infractions or norm violations (Chudek and Henrich, 2011; Wiessner, 2005). How human morality evolved therefore depends on understanding how an originally primate-style dominance hierarchy, headed by the best fighter, transitioned

into a human-style moral system, headed by a dominant alliance within which there is a norm of egalitarianism (Boehm, 2012).

2. In great ape groups, as is typical of non-human primates, physical contests among males leads to one male becoming the highest-ranked fighter, or alpha. Alpha males use their domination of other males to achieve disproportionately high reproductive success (*Pan*: McCarthy et al., 2020; *Gorilla*: Robbins et al., 2014; *Pongo*: Tajima et al., 2018; other primates: Minkner et al., 2018). This is true even in bonobos *Pan paniscus*, despite the alpha male typically being co-dominant with the alpha female and depending on his mother for support (Surbeck et al., 2019). Sexual dimorphism in the body mass of early *Homo* was at least as high as in living humans, indicating that male fighting ability consistently exceeded that of females (Villmoare et al., 2019). Thus based on ape behavior and comparative anatomy, the dominance hierarchy of any recent ancestor of humans prior to the development of a moral system can be reconstructed as having been headed by an alpha male.

3. Human males are hardly ever alphas in the primate sense. Kings, emperors, presidents etc are sometimes informally called alpha males, but those individuals differ from non-human alphas because they do not earn their leadership by personally fighting all challengers. Instead, they are leaders of alliances; their fate depends on the fighting prowess of their alliance, and on whether their alliance continues to support them. This is why the term "reverse dominance hierarchy" is an apt description of the human system: every member of the alliance is subject to being dominated by his allies (Boehm, 1993).[1] Contexts in which genuine alphas might be found in humans, such that the most dominant individuals are those who personally fight all challengers, are likely restricted to small groups such as children's play-groups or gangs of adolescents.

4. The time when a pre-moral system of conflict regulation transitioned to a moral system should in theory be identifiable from a reduction in the intensity of selection for alpha-male behavior. In the pre-moral era, alpha males would have used reactive (impulsive) aggression to respond rapidly and violently to perceived status challenges, as nonhuman primates do (e.g. Goodall, 1986). Evidence of selection against reactive aggression is therefore

[1] The term "reverse dominance hierarchy" can be criticized on the basis that in humans, the dominance relationship between a potential tyrant and an alliance of his subordinates is actually not reversed: it is equalized. Boehm (1993) justified his use of the term, however, by noting that a potential alpha is vulnerable to being killed by the egalitarian group.

expected to signal a shift towards selection against alpha-style behavior. Such evidence is expected to be visible because in domesticated animals, selection against reactive aggression leads to the emergence of a "domestication syndrome", i.e. a characteristic series of anatomical, behavioural, physiological and cognitive traits. *H. sapiens* exhibit multiple anatomical elements of the domestication syndrome compared to *H. heidelbergensis*, including reductions in body mass, brain size and sexual dimorphism, shorter and narrower face, smaller molars and reduced trabecular bone compared to earlier ancestors (Leach, 2003; Cieri et al., 2014; Wrangham 2021). Reductions in two of these traits (face and molars) are first detectable from ~315,000 years ago at Jebel Irhoud in Morocco, which justifies those fossils being named the earliest *H. sapiens* (Hublin et al., 2017). Other traits develop *sapiens*-style features at various subsequent times. The initial detection of *sapiens*-style features at ~300,000 years ago thus indicates that by then our ancestors were beginning to lose the alpha-male system.

5. Preliminary genetic evidence supports the scenario of a self-domestication event that started with *H. sapiens*. When compared to their wild ancestors, domesticated and self-domesticated species display certain genetic changes in parallel related to features of the domestication syndrome. Although no genetic material is yet available from *H. heidelbergensis*, two species of *Homo* that lived contemporaneously with *H. sapiens* prior to 40,000 years ago offer useful stand-ins for *H. heidelbergensis*: they are *H. neandertalensis* (in Europe and southwest Asia) and *H. "denisova"* (in western Asia). The lineage leading to *H. neandertalensis* and *H. "denisova"* split from *H. sapiens*' ancestors around 500,000 years ago, and neither of those relatives shows the anatomical signs of self-domestication found in *H. sapiens*. Genetic changes associated with mammalian domestication have now been found in *H. sapiens* compared to *H. neandertalensis* and *H. "denisova"* (Theofanopolou et al., 2017; Zanella et al., 2019). In one analysis the time when such genes were initially favoured in *H. sapiens* was narrowed to the expected time, i.e. 500,000 to 300,000 years ago (Andirko et al., 2021). Genomes thus offer *a priori* tests of the hypothesis that *H. sapiens* is a self-domesticated form; and initial tests are supportive.

6. The new social dynamic that is indicated at ~300,000 years ago appears to have been maintained ever since. Evolutionary changes from Jebel Irhoud to the present indicate a continuing reduction in the propensity for reactive aggression. For example facial width has fallen throughout the existence of *H. sapiens*, and is associated with reactive aggression. Within contemporary

populations of Americans, Europeans and Chinese, men with relatively narrower faces tend to be less reactively aggressive, as well as being perceived as being less threatening (Short et al., 2012; Geniole et al., 2015). Overall, the anatomical evidence indicates that there has been a continuing selective pressure against reactive aggression, and therefore against alpha-style behaviour, for a little more than 300,000 years.

7. The novel selection pressure that first acted against reactive aggression in *Homo heidelbergensis* is most likely to have been social. This inference comes from the fact that, as happens in other primates, a male hunter-gatherer could in theory obtain genetic dividends by becoming a primate-style alpha who obtained mating success by using physical force against rival males and/or fertile females. Among hunter-gatherers such attempts do occasionally occur (Boehm, 2012). Even when outright conflict does not occur, the threat of its happening is still present: among hunter-gatherers "the dangers of conflict between men over claims not only to women but more generally to wealth, to power or to prestige are well understood" (Woodburn, 1982, p. 436).
8. Capital punishment is known worldwide, including among hunter-gatherers in every continent. When a male hunter-gatherer uses tyrannical behaviour to attempt to dominate others, initial efforts to control him consist of non-violent tactics such as public criticism, ridicule, derisory singing or ostracism. When such efforts fail, communities that have no police or prisons eventually resort to capital punishment. The execution can be planned in advance or justified later. It can be conducted by any number of people from a lone adult to a united group. It is overwhelmingly carried out by men (Boehm 1999, 2012, 2017, 2018).
9. Proactive (premeditated) and reactive aggression are controlled in animals by partially separate neural pathways, and the same appears true of humans (Wrangham, 2018). This means that when a Pleistocene execution (which would have used proactive aggression) caused the death of a domineering bully (who had a high propensity for reactive aggression), the typical pattern was for reactive aggression to be selected against. In small-scale societies capital punishment is the only mechanism known to systematically operate against violent domineering behaviour. Evidence of the domestication syndrome in the fossil record, as seen in the origin of *H. sapiens*, is therefore readily attributable to the evolution of capital punishment. Other proposed mechanisms, such as female choice of less aggressive males as mating

partners, fail to explain how the bullying behaviour of a determined and domineeering male would be constrained (Wrangham, 2019).

10. In independent small-scale societies today, the legitimate use of violence is monopolized by an alliance of elders who organize or permit executions conducted in defence of social norms. According to the execution hypothesis, essentially the same monopoly on the legitimate use of violence has been maintained throughout the >300,000-year existence of *Homo sapiens*, elaborated nowadays by institutions including law and religion. This inference suggests that on the one hand, the ultimate source of ethical concepts such as justice is the set of norms created or approved by such a male alliance; and on the other hand, the ultimate source of individuals' readiness to conform to such concepts is the selective pressure of capital punishment that killed non-conformists throughout the last 300,000 to 400,000 years. Human social-psychological tendencies have thus evolved to be deeply, albeit often subconsciously, self-protective by promoting conformity to group norms. The effectiveness of recent non-lethal forms of punishment such as prisons, castration, fines and exile has reduced the frequency of execution, but Pleistocene principles still hold: non-conformists are punished in ways that stop them from undermining the interests of the male alliance, and the incidental effect is that the genetic fitness of the non-conformists tends to be reduced, leading to selection against antisocial behavior.
11. Human moral intuitions tend to protect agents from accusations of immorality by pushing agents towards plausibly deniable and/or conforming actions (Wrangham, 2019). For example the "Action/Omission" bias describes individuals preferring omission to commission. The "Means/Side-Effect" bias pushes agents from actions that intentionally lead to harm. The "Contact/Non-contact" bias reflects efforts to avoid touching someone who is being harmed. A bias for conformity means that moral decisions are strongly influenced by social context (Kundu and Cummins, 2013), even when "social context" is as limited as the presence of two black spots that resemble eyes (Engelmann et al., 2016). The tendency for morally guided actions to be conformist and self-protective fits the proposal that social punishment has been a key component of the evolution of the moral senses, leading to a morality that serves the prevailing ethics of the group, or a power-wielding subgroup, rather than being derived from any rationally derived absolute (Paley, 2021).

In sum, the execution hypothesis is inspired by a broad range of biological and ethnographic data, some of which now serve as *a priori* tests.

## ALTERNATIVE APPROACHES.

Moral behaviour is widely agreed to reduce competition and conflict, and therefore to benefit group members on average (Burkart et al., 2017; Curry et al., 2019a). The execution hypothesis proposes that this positive relationship between morality and benefits began as an incidental consequence of the way that group males resolved conflicts among themselves: males first developed a newly sophisticated style of cooperation to kill alphas, and then adapted that skill towards promoting a broader set of their interests. By contrast, a longer tradition sees the benefits of group-level cooperation as being the primary reasons why morality evolved. Various potential benefits of cooperation have been proposed, such as increased food-sharing within groups (Tomasello, 2016) or increased resource sharing between groups (Spikins et al., 2021), but the most frequently proposed candidate is an increased effectiveness in intergroup competition (Choi and Bowles, 2007).

A leading example of a theory explaining morality at least partly as an adaptation for increasing group competitive ability is the ecological dominance social competition model (EDSC) developed by Alexander (1971, 1979, 1982, 1987, 1990) (Flinn et al., 2005; Summers et al., 2020). Alexander suggested that during the Pleistocene, *Homo* groups had overcome the hostile forces of nature sufficiently well that social competition in general, and warfare in particular, became predominant selective forces on their behaviour. Under these conditions selection favoured groups whose members were the most moral, and who therefore became the most cooperative and successful in intergroup competition.

In support of this premise, warfare was likely an important selective pressure in the Pleistocene (Glowacki et al., 2020). Numerous experiments and observations show that intergroup competition tends to increase within-group cooperation in humans (Bauer et al., 2016, Henrich and Muthukrishna, 2021), as well as in other species (chimpanzees: Brooks et al., 2021, Samuni et al., 2020; birds and mammals: Radford et al., 2016). Within-group cooperation can also increase the likelihood of winning intergroup conflicts (Turchin et al., 2013).

In many ways the EDSC and the execution hypothesis are reconcilable. Alexander (1982) proposed that moral systems developed when individuals gained the cognitive ability to manipulate other group members into cooperating, including in ways that would aid their group in intergroup competition. The conferring of rewards and punishments was critical. Systems of indirect reciprocity and monitoring of reputations

meant that beneficent acts given by an agent to individual A could lead to the agent receiving return benefits at later times from individuals B, C etc. Within such constraints Alexander viewed individuals as constantly striving to maximise net gain. For example whereas Darwin (1871) conceived of a conscience as a device to inhibit immoral behavior, Alexander (1979) saw it as a mechanism for strategizing how morally an agent should behave in a given situation. In Alexander's view selection favoured individuals who gave the appearance of acting for the greater good, even if the appearance was false. Thus an essential feature of Alexander's argument was that morality evolved in response to a tension between the interests of each individual and their group. These concepts are broadly compatible with the execution hypothesis, as Boehm (2012) emphasized in acknowledging Alexander's contributions.

Against the EDSC, however, it does not address the emergence of the reverse dominance hierarchy, it is silent on the question of specifically when morality emerged, and it has not been related to the paleo-anthropological record, including the question of when warfare supposedly became increasingly important. It also faces specific difficulties.

First, an appealing feature of the EDSC model, as it was first conceived, was that it suggested that a unique feature (human morality) was explicable by a unique evolutionary stimulus (exceptionally intense warfare). Subsequent research on intergroup aggression found, however, that chimpanzees similar death rates from intergroup aggression as hunter-gatherers (Wrangham et al., 2006). This suggests that although intergroup aggression was important for Pleistocene *Homo* its selective significance was not exceptional.

Second, mechanisms by which morality could have evolved through its effects on intergroup competition in the Pleistocene have not yet been identified (Dyble, 2021). The essential problem is that the free-riders who do not cooperate (do not self-sacrifice) in war would apparently benefit from the self-sacrificial efforts of their peers. Theories of group-structured cultural evolution readily attribute the growing scale of human cooperation in the last 12,000 years to the effects of warfare in promoting group norms (Henrich and Muthukrishna, 2021), but whether such models are applicable to earlier phases of human evolution has not been demonstrated.

In sum, Alexander's theory and the execution hypothesis share the view that morality evolved from social selection of behaviour, with individuals trying to maximise their gain within societies of intelligent, problem-solving, cooperative and competitive egoists. Alexander would likely have agreed with the execution hypothesis in its answer to the normative question considered in this paper, i.e. "What is the logic for an agent feeling s/he has to perform a given moral act?" According to both approaches, the proximate answer would be that the agent should follow the moral rule unless s/he can

be confident of escaping the potential costs of a more selfish act; and the ultimate answer is that the agent's moral emotions have evolved under social selection in ways that tend to benefit the agent when in the public eye by conforming.

On the other hand, the hypothesis that cooperation for warfare underlies the evolution of morality is weakened by its being untied to the paleo-anthropological record, its silence on the question of how alpha-style violence was selected against, its error in predicting that intergroup aggression would be uniquely important for humans compared to other species, and its lack of a convincing mechanism. No other group-benefit theories are markedly more successful than Alexander's in identifying the social dynamics by which moral behaviour would have been favoured (Dyble, 2021). There is certainly strong evidence that moral rules tend to promote cooperative behaviour and that very similar moral rules are widespread or universal across societies of all types (Curry et al., 2019a, 2019b). But the fact that morality promotes cooperation, whether in food production, food-sharing, caring, war or other contexts, does not explain why and how it emerged.

## SEX DIFFERENCES IN THE ORIGIN AND DYNAMICS OF THE MORAL SYSTEM.

Korsgaard (1996) apparently follows a long tradition in moral philosophy of treating the moral system as having emerged from cognitive processes for which sex differences are irrelevant. The question of why moral attitudes often favour males over females might then be answered by the idea that males respond to a system that was originally unbiased by sex, and exploit it to benefit themselves.

In a somewhat parallel way, evolutionary theorists attempting to explain why humans are an "ultrasocial" (highly cooperative) species routinely do so without addressing sex differences in cooperative tendencies, or the presence of major sex biases in institutions that support cooperation. For instance in a major recent review of the topic Henrich and Muthukrishna (2021) did not mention sex differences, while at the same time stressing that the interplay among cultural institutions, social norms and cooperative psychology represents a vital contribution to making humans more cooperative than other species. Again, therefore, the implication is that patriarchal inclinations are secondary developments from a system of elaborated cooperation that evolved in the species as a whole.

The execution hypothesis, by contrast, proposes that from the outset, the driving force for the evolution of morality was the behaviour of adult males. An alpha male's domineering aggression is directed mainly towards his male rivals, not to females. Subordinate males will therefore normally tend to benefit by engineering the removal

of an alpha. In typical mammals, the effect is merely to increase the chance that a subordinate male will become the new alpha whereas in linguistically skilled *Homo*, the effect is much better for the majority of subordinates (i.e. for those who would not have replaced the alpha): it means that the alpha position is abolished, and the subordinates become members of the dominant alliance.

In contrast to the predictable benefits to males, how much females would stand to gain from the removal of an alpha male is questionable. The answer should depend partly on how much aggression the females would have received from the alpha, which is unknown. An alpha *Homo* would not necessarily have behaved towards females in a domineering manner. For instance, gorillas are a species in which the alpha male is about twice the weight of females and entirely dominant to them. Yet the alpha gorilla's aggression towards females is mostly mild to moderate, and sexual coercion is minimal (Palombit, 2014). By contrast, in groups in which females mate with multiple males the rates of aggression from males towards females can be high (Goodall, 1986). In short, with respect to the amount of aggression that they received it is not clear whether female *Homo heidelbergensis* would have benefited by the change from a male hierarchy headed by an alpha to one headed by an alliance.

In line with the Pleistocene scenario, two of the main institutions that support moral norms in the ethnographic present are overwhelmingly dominated by males, namely law and religion. In every kind of society from small-scale to state, females can participate importantly in male-dominated institutions, but rarely or never as a majority. Unsurprisingly, therefore, the social systems whose moral nature those institutions forge are consistently patriarchal: male interests tend to be prioritized on such diverse questions as who is allowed to eat the best food, what punishments should be given for violation of sexual norms, who inherits what, or who is allowed to punish who (Rosaldo and Lamphere, 1974; Smuts, 1992, 1995; Hudson et al., 2020).

What explains these patriarchal tendencies of contemporary *H. sapiens*? A theoretical possibility is that the patriarchal nature of institutional morality is an example of evolutionary inertia. In other words, it could be a formerly adaptive system that is now, due to changed circumstance, non-adaptive but maintained because of psychological tendencies that are no longer well matched to current circumstance.

Alternatively if the core rationale for the evolution of morality was to constrain the domineering aggression of selfish males, the same ancient logic might still apply. Male reactive aggression has apparently been continuously selected against for at least 300,000 years, but even now, in every society, male violence is still a problem that has to be managed, and is much more of a social problem than female violence. The males who do the threatening have changed from the middle Pleistocene. They are no longer merely tyrants acting in the style of an alpha *H. heidelbergensis* or great ape; now they

can also be members of a threatening coalition. But either way, perpetrators of violence must be confronted and constrained. Even though in many ways females suffer worse from male violence than males do, the individuals who have most to gain from stopping males from being violent are arguably other males, since they, not females, are the competitors for genetic fitness. This suggests that even today, patriarchal aspects of the moral system are driven by male efforts to compete for power with other males of their own group.

## THE PERSISTENT IMPORTANCE OF MORAL ENFORCEMENT

According to Korsgaard (1996, p. 8), Hobbes (1651) claimed that there is no right or wrong in the state of nature. If Hobbes had been referring to non-humans, he would have been correct. But if his "state of nature" was intended to include humans living as hunter-gatherers, he could hardly have been more mistaken. Ever since Durkheim (1902), hunter-gatherers and others living in small-scale, acephalous bands have been known to live by a set of norms that categorize numerous behaviours as right or wrong. Morally circumscribed behaviors concern food, sharing, sexuality, marriage partners, emotional expression, disrespect, secret societies and much else, and are the topic of much daily conversation. To judge from one detailed study of Ju/'hoansi Bushmen hunter-gatherers, moral enforcement comes more from punishment than reward, with males being sanctioned more than females (Wiessner, 2005).

Intense forms of punishment, as I have argued, can explain why human moral emotions have evolved to be intensely self-protective. Does this mean that the reason why contemporary people are virtuous is that they fear punishment? Korsgaard (1996) was sceptical. In response to Mandeville (1714) claiming that "virtue is just an invention of politicians, used to keep their human cattle in line" (Korsgaard, 1996, p. 8), she said that if that were true, ordinary people would have insufficient reason to be virtuous. The implication, apparently, was that ordinary people experience little moral enforcement. Perhaps she was right with respect to the influence of Mandeville's politicians being distant and minor, but she was surely wrong if she thought that individuals in stable human communities are not affected by the moralistic aggression of their peers. Morality acts locally in the gossip and disapproval of peers, leading easily to direct confrontations, ostracism or violence. Those sanctions seem to have been as important in determining how the moral emotions evolved as they continue to be in shaping how we behave today.

What logic, then, explains to an observer why an honest agent feels that s/he has to follow a given moral rule? The execution hypothesis applies to emotional responses rather than explicit rationalizations. It proposes that one reason why an agent feels that

she has to follow a given moral rule is that she is descended from more than 12,000 generations in each of which there was a social alliance ready to punish those whose behaviour did not conform to male interests. An inadvertent result was to create genetic selection in favour of those whose emotional systems were primed to conform to the learned norms of her community. There need be nothing conscious or contextually explicable in the agent's current response therefore. Her moral decision comes from an unintended legacy, bequeathed by the self-protective behavior of both sexes in response to the selfish threats of males competing for power in a uniquely human style.

## ACKNOWLEDGMENTS.

I thank Chris Boehm and Brian Hare for longterm collaboration, Diane Rosenfeld for discussion, Barb Smuts for comments on the reverse dominance hierarchy, an anonymous reviewer for suggestions, and Bernie Crespi, Mark Flinn and Kyle Summers for advice about Richard Alexander's thinking. Thanks also to Roberto Redaelli for the invitation to participate in the conference on 'Rethinking the Sources of Normativity in Ethics' (Friedrich-Alexander-Universität Erlangen-Nürnberg, March 2021).

## BIBLIOGRAPHY.

Alexander, R. D. (1971). The search for an evolutionary philosophy of man. *Proc Roy Soc Victoria, 84*, 99-120.

Alexander, R. D. (1979). *Darwinism and Human Affairs*. Seattle WA: University of Washington Press.

Alexander, R. D. (1982). Biology and the moral paradoxes. *J. Social Biol. Struct., 5*, 389-395.

Alexander, R. D. (1987). *The Biology of Moral Systems*. New York: Aldine de Gruyter.

Alexander, R. D. (1990). *How did humans evolve? Reflections on the uniquely unique species.* Ann Arbor, MI: The University of Michigan.

Andirkó, A., Moriano, J., Vitriolo, A., Kuhlwilm, M., Testa, G., & Boeckx, C. (2021). Fine-grained temporal mapping of derived high-frequency variants supports the mosaic nature of the evolution of *Homo sapiens*. *bioRxiv 2021.01.22.427608*. doi: 10.1101/2021.01.22.427608

Andrews, K. (2020). Naive normativity: the social foundation of moral cognition. *Journal of the American Philosophical Association, 6*(1), 36-56. doi:10.1017/apa.2019.30

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., & Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives, 30*, 249-274. doi: 10.1257/jep.30.3.249

Boehm, C. (1993). Egalitarian behavior and reverse dominance hierarchy. *Current Anthropology, 34*(3), 227-240.

Boehm, C. (1999). *Hierarchy In The Forest: The Evolution Of Egalitarian Behavior.* Cambridge, MA: Harvard University Press.

Boehm, C. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. New York: Basic Books.

Boehm, C. (2017). Ancestral precursors, social control, and social selection in the evolution of morals. In M. N. Muller, R. W. Wrangham, & D. P. Pilbeam (Eds.), *Chimpanzees and Human Evolution* (pp. 746-790). Cambridge, MA: Harvard University Press.

Boehm, C. (2018). Collective intentionality: A basic and early component of moral evolution. *Philosophical Psychology, 31*(5), 680-702. doi: 10.1080/09515089.2018.1486607

Brooks, J., Onishi, E., Clark, I. R., Bohn, M., & Yamamoto, S. (2021). Uniting against a common enemy: Perceived outgroup threat elicits ingroup cohesion in chimpanzees. *PLoS ONE, 16*(e0246869), 1-17. doi: 10.1371/journal. pone.0246869

Brosnan, S. F., & de Waal, F. B. M. (2014). Evolution of responses to (un)fairness. *Science, 346*, 1251776. doi:10.1126/science.1251776

Burkart, J. M., Brügger, R. K., & van Schaik, C. P. (2018). Evolutionary origins of morality: insights from non-human primates. *Frontiers in Sociology, 3*(17), 1-12. doi:10.3389/fsoc.2018.00017

Choi, J.-K., & Bowles, S. (2007). The coevolution of parochial altruism and war. *Science, 318*, 636-640.

Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences, 15,* (5), 218-226.

Cieri, R. L., Churchill, S. E., Franciscus, R. G., Tan, J., & Hare, B. (2014). Craniofacial feminization, social tolerance, and the origins of behavioral modernity. *Current Anthropology, 55*, 419-443.

Curry, O. S., Chesters, M. J., & Van Lissa, C. J. (2019b). Mapping morality with a compass: Testing the theory of 'morality-as-cooperation' with a new questionnaire. *Journal of Research in Personality, 78*, 106-124. doi: 10.1016/j.jrp.2018.10.008

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019a). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology, 60*(1), 47-69.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: J. Murray.

de Waal, F. B. M. (2006). *Primates and Philosophers: How Morality Evolved*. Princeton NJ: Princeton University Press.

Durkheim, É. (1902). *The Division Of Labor In Society*. New York: Free Press.

Dyble, M. (2021). The evolution of altruism through war is highly sensitive to population structure and to civilian and fighter mortality. *Proceedings of the National Academy of Sciences, 118*(e2011142118), 1-6. doi: 10.1073/pnas.2011142118

Engelmann, J. M., Clift, J. B., Herrmann, E., & Tomasello, M. (2017). Social disappointment explains chimpanzees' behaviour in the inequity aversion task. *Proceedings of the Royal Society B, 284*, 20171502. doi: 10.1098/rspb.2017.1502

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2016). The effects of being watched on resource acquisition in chimpanzees and human children. *Animal Cognition, 19*(1), 147-151.

Flanagan, J. G. (1989). Hierarchy in simple "egalitarian" societies. *Annual Review of Anthropology, 18*, 245-266.

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior, 26*, 10–46. doi:10.1016/j.evolhumbehav.2004.08.005

Geniole, S. N., Denson, T. F., Dixson, B. J., Carré, J. M., & McCormick, C. M. (2015). Evidence from meta-analyses of the facial width-to-height ratio as an evolved cue of threat. *PLoS ONE 10*(7), e0132726. doi:10.1371/journal.pone.0132726

Gintis, H., van Schaik, C., & Boehm, C. (2015). Zoon Politikon: the evolutionary origins of human political systems. *Current Anthropology, 56*(3), 327-353.

Glowacki, L., Wilson, M. L., & Wrangham, R. (2020). The evolutionary anthropology of war. *Journal of Economic Behavior and Organization, 178*, 963-982. doi:10.1016/j.jebo.2017.09.014

Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Cambridge MA: Harvard University Press.

Greene, J. D., & Young, L. (2020). The cognitive neuroscience of moral judgment and decision-making. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (Vol. 6). Cambridge, MA: MIT Press.

Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Pantheon.

Hare, B. (2017). Survival of the friendliest: *Homo sapiens* evolved via selection for prosociality. *Annual Review of Psychology, 68*, 155-186.

Henrich, J., & Muthukrishna, M. (2021). The origins and psychology of human cooperation *Annual Review of Psychology, 72*, 207-240.

Hobbes, T. (1651). *Leviathan* (R. Tuck Ed.). Cambridge, England: Cambridge University Press.

Hublin, J.-J., Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., . . . Gunz, P. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. *Nature, 546*, 289-292.

Hudson, V. M., Bowen, D. L., & Nielsen, P. L. (2020). *The First Political Order: How Sex Shapes Governance and National Security Worldwide*. New York: Columbia University Press.

Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.

Kundu, P., & Cummins, D. D. (2013). Morality and conformity: The Asch paradigm applied to moral decisions. *Social Influence, 8*(4), 268-279. doi:10.1080/15534510.2012.727767

Leach, H. (2003). Human domestication reconsidered. *Current Anthropology, 44*(3), 349-368.

Mandeville, B. (1714). *The Fable of the Bees: or, Private Vices, Public Benefits* (F. B. Kaye Ed.). Indianapolis: Liberty Classics.

McAuliffe, K., & Santos, L. R. (2018). Do animals have a sense of fairness? In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology* (pp. 393-400). NY: Guilford Press.

McCarthy, M. S., Lester, J. D., Cibot, M., Vigilant, L., & McLennan, M. R. (2020). Atypically high reproductive skew in a small wild chimpanzee community in a human-dominated landscape. *Folia primatologica, 91*, 688-696. doi:10.1159/000508609

Minkner, M. M. I., Young, C., Amici, F., McFarland, R., Barrett, L., Grobler, J. P., . . . Widdig, A. (2018). Assessment of male reproductive skew via highly polymorphic STR markers in wild vervet monkeys, *Chlorocebus pygerythrus. Journal of Heredity, 2018*, 780–790. doi:10.1093/jhered/esy048

Paley, C. (2021). *Beyond Bad: How Obsolete Morals Are Holding Us Back.* London: Hodder & Stoughton.

Palombit, R. A. (2014). Sexual conflict in nonhuman primates. *Advances in the Study of Behavior, 46*, 191-280.

Radford, A. N., Majolo, B., & Aureli, F. (2016). Within-group behavioural consequences of between-group conflict: a prospective review. *Proc R Soc B, 283*(20161567), 1-10. doi:10.1098/rspb.2016.1567

Robbins, A. M., Gray, M., Uwingeli, P., Mburanumwe, I., Kagoda, E., & Robbins, M. M. (2014). Variance in the reproductive success of dominant male mountain gorillas. *Primates, 55*(4), 489-499. doi:10.1007/s10329-014-0426-2

Rosaldo, M. Z. (1974). Women, culture and society: a theoretical overview. In M. Z. Rosaldo & L. Lamphere (Eds.), *Woman, Culture and Society* (pp. 17-42). Stanford, CA: Stanford University Press.

Samuni, L., Mielke, A., Preis, A., Crockford, C., & Wittig, R. M. (2020). Intergroup competition enhances chimpanzee (*Pan troglodytes verus*) in-group cohesion. *International Journal of Primatology, 41*, 342–362. doi:10.1007/s10764-019-00112-y

Short, L. A., Mondloch, C. J., McCormick, C. M., Carré, J. M., Ma, R., Fu, G., & Lee, K. (2012). Detection of propensity for aggression based on facial structure irrespective of face race. *Evolution and Human Behavior, 33*, 121–129.

Smuts, B. B. (1992). Male aggression against women: an evolutionary perspective. *Human Nature, 3*, 1-44.

Smuts, B. B. (1995). The evolutionary origins of patriarchy. *Human Nature, 6*, 1-32.

Sripada, C. S., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *Evolution and Cognition. The Innate Mind, Vol. 2. Culture and Cognition* (pp. 280-301). Oxford: Oxford University Press.

Summers, K., Crespi, B. J., & Flinn, M. V. (2020). Were humans their own most important selective pressure for cooperation and morality? A critical review of Richard Wrangham's *The Goodness Paradox*. *Evolutionary Behavioral Sciences*. doi:10.1037/ebs0000203

Surbeck, M., Boesch, C., Furuichi, T., Fruth, B., Hohmann, G., Ishikuza, S., . . . Langergraber, K. (2019). Males with a mother living in their community have higher reproductive success in bonobos but not chimpanzees. *Current Biology, 29*, R1–R3.

Tajima, T., Malim, T. P., & Inoue, E. (2018). Reproductive success of two male morphs in a free-ranging population of Bornean orangutans. *Primates, 59*, 127-133. doi:10.1007/s10329-017-0648-1

Theofanopoulou, C., Gastaldon, S., O'Rourke, T., Samuels, B. D., Martins, P. T., Delogu, F., . . . Boeckx, C. (2017). Self-domestication in Homo sapiens: Insights from comparative genomics. *PLoS ONE, 12*(10), e0185306. doi: 10.1371/journal.pone.0185306

Tomasello, M. (2016). *A Natural History of Human Morality*. Cambridge MA: Harvard University Press.

Turchin, P., Currie, T. E., Turner, E. A. L., & Gavrilets, S. (2013). War, space, and the evolution of Old World complex societies. *Proceedings of the National Academy of Sciences, 110*(41), 16384-16389. doi:10.1073/pnas.1308825110

Villmoare, B., Hatala, K. G., & Jungers, W. (2019). Sexual dimorphism in *Homo erectus* inferred from 1.5Ma footprints near Ileret, Kenya. *Scientific Reports, 9*(7687), 1-12. doi:10.1038/s41598-019-44060-2

Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi Bushmen: a case of strong reciprocity? *Human Nature, 16*(2), 115-145.

Woodburn, J. (1982). Egalitarian societies. *Man, 17*(3), 431–451.

Wrangham, R. W. (2018). Two types of aggression in human evolution. *Proceedings of the National Academy of Sciences, 115*(2), 245-253. doi: 10.1073/pnas.1713611115

Wrangham, R. W. (2019). *The Goodness Paradox: The Strange Relationship Between Virtue and Violence in Human Evolution*. New York: Alfred A. Knopf.

Wrangham, R. W. (2021). Targeted conspiratorial killing, human self-domestication and the evolution of groupishness. *Evolutionary Human Sciences, 3*(e26), 1-21. doi:10.1017/ehs.2021.20

Wrangham, R. W., Wilson, M. L., & Muller, M. N. (2006). Comparative rates of aggression in chimpanzees and humans. *Primates, 47*, 14-26.

Zanella, M., Vitriolo, A., Andirko, A., Martins, P. T., Sturm, S., O'Rourke, T., . . . Testa, G. (2019). Dosage analysis of the 7q11.23 Williams region 1 identifies BAZ1B as a major human gene patterning the modern human face and underlying self-domestication. *Science Advances, 5* (eaaw7908).