eISSN: 1989-3612
DOI: https://doi.org/10.14201/art2024.31231

EVOLVED AND CULTURAL INTUITIONS. HIGHLY SPECULATIVE REMARKS ON THE ORIGINS OF OUR SENSE OF FAIRNESS

Intuiciones evolutivas y culturales. Comentarios fuertemente especulativos sobre los orígenes de nuestro sentido de justicia

Rodrigo BRAICOVICH
Universidad Nacional de Rosario, CONICET
rbraicovich@gmail.com

Recibido: 6/03/2023 Revisado: 17/04/2023 Aceptado: 23/06/2023

ABSTRACT: The question of whether the sense of fairness constitutes an exclusively human trait has been answered mostly from two polar positions: the first one unambiguously affirms such exclusivity, thus denying the relevance of cognitive ethology to understand our evaluations of justice; the second one, on the contrary, postulates the existence of a (proto) sense of fairness in non-human animals, strongly related to ours, which would make cognitive ethology highly relevant to understand the mechanisms on which our evaluative practices are based. From a position of extreme caution in relation to the possibility of (eventually) offering concrete evidence in favor of innatist theses such as the one I will defend here, I will suggest that i) in line with the rupturist positions, it is possible to preserve the human exclusivity of the sense of justice, ii) in line with the continuist positions, the relevance of studies coming from cognitive ethology is guaranteed, insofar as (*ex hypotesi*) our evaluative practices often

take as input innate psychological dispositions shared with other species. Finally, I will suggest that the concept of rationalization is central to determine in each case the possible articulation between innate dispositions and explicit justifications.

Keywords: fairness evaluations, moral psychology, phylogenesis, justice, rationalization.

RESUMEN: La pregunta acerca de si el sentido de justicia constituye un rasgo exclusivamente humano ha sido respondida mayormente desde dos posiciones polares: la primera de ellas afirma sin ambages dicha exclusividad, negando con ello la relevancia de la etología cognitiva para comprender nuestras evaluaciones de justicia; la segunda, por el contrario, postula la existencia de un (proto) sentido de justicia en animales no humanos, fuertemente emparentado con el nuestro, lo cual volvería a la etología cognitiva sumamente relevante para comprender los mecanismos sobre los que se basan nuestras prácticas evaluativas. Desde una postura de extrema cautela en relación con la posibilidad de ofrecer (eventualmente) evidencia concreta en favor de tesis innatistas como la que defenderé aquí, sugeriré que i) en línea con las posturas rupturistas, es posible preservar la exclusividad humana del sentido de justicia, atendiendo al hecho de que nuestras prácticas evaluativas presuponen el uso de conceptos, y que ii) en línea con las posturas continuistas, la relevancia de los estudios provenientes de la etología cognitiva se halla garantizada, en la medida en que (ex hypotesi), nuestras prácticas evaluativas frecuentemente toma como *input* disposiciones psicológicas innatas compartidas con otras especies. Sugeriré, por último, que el concepto de racionalización es clave para determinar en cada caso la posible articulación entre disposiciones innatas y justificaciones explícitas.

Palabras clave: evaluaciones de justicia, psicología moral, filogénesis, justicia, racionalización.

1. Introduction

The idea that we are born with certain dispositions that are shared by every human being, or that there is, in other words, such a thing as a human nature, has been debated in Western thought since at least the Hellenistic period. The idea that there are patterns of behavior that derive from an innate makeup has been resisted fiercely on account both of its methodological flaws, its political underpinnings and its ethical consequences (see, for instance, Dupré, 2001; Gould, 1996; Levins & Lewontin, 2009; Sahlins, 2008), but the very denial of human nature has, in turn, been

met with a similar reaction, since it has been contested not on methodological grounds but mainly because of the alleged political agenda that has driven such a denial (see Haidt, 2012; Mosterín, 2011; Pinker, 2002; Wilson, 1997). Contrary to what the news portals and the media might suggest with their permanent insistence on every alleged discovery of human traits that might be present in other species or that have somehow been proven to be universal, the heyday of the investigations concerning human nature is long over. After the first wave of sociobiology had passed, the apparently accelerated and exponential developments in genetics, neurology, developmental psychology, cognitive ethology and biological anthropology revived the promise that, led by the hand of evolutionary psychology, a new and definitive comprehension of the evolutionary origins of human nature would be possible. Things happened in the way, unfortunately: both the modularity thesis and the basic emotions theory (two not so often acknowledged pillars of the models of human nature en vogue by the end of the 20th century) came under heavy fire, paving the way for the resurgence of cultural and constructionist approaches (see Brooks et al., 2017; Buller & Hardcastle, 2000; Feldman Barrett, 2017; Feldman Barrett & Russell, 2015; Fox & Friston, 2012; George & Sunny, 2019; Palecek, 2017; Prinz, 2006; Suhler & Churchland, 2011). The quest for human nature seems thus to be facing its most serious crisis since Europe discovered the American and African Other, and the prospect of understanding the evolutionary basis on which some of the (allegedly) most recurrent traits in humans seems to dwindle as the methodological objections to such an enterprise keep piling up.

Does this mean that we should abandon all hope of ever settling the question concerning whether there is a substantive and shared ground that explains at least part of what constitutes us as humans? Are we forced to leave behind the idea that there is a part of us that is shared by non human animals and that by understanding them we may better understand ourselves – at least in what matters from a psychological, ethical and social perspective? Does ethology, in other words, have nothing to offer to philosophical anthropology? I believe that it is not necessary to arrive to such conclusions. But I believe as firmly that any conclusions we may wish to arrive at concerning the question of the existence of human nature, its evolutionary origins and the possible contributions of ethology to philosophical anthropology must remain extremely cautious in its formulation. All expressions must be preceded by "might" and nuanced by "perhaps". And if we cannot rule out the possibility that a day may come when we may perhaps be forced to abandon such hopes, we can at least continue to explore the most plausible pathways that are still open

to pursue. One of them is the one that was opened by Frans de Waal's pioneering works on the existence of a (proto) sense of fairness in non human animals, and the aim of the following pages will be to ask whether it is a path that we can still fruitfully explore.

2. Sense of fairness: human and non human

The idea that our sense of fairness precedes our species and can be traced to other non human animals with which we share a common evolutionary past is not new: since Charles Darwin suggested in *The Descent of Man* (1871/1880) the presence of (what we call) moral emotions in several species other than ours, the evidence in favor of the hypothesis of a (proto) sense of justice in other animals seemed to accumulate considerably. De Waal and Sarah Brosnan, on the one hand, and Mark Bekoff and Jessica Pierce, on the other, were among the most prominent defenders of that idea. Since as far as 1991, De Waal has defended (with variations) the idea that "the human sense of justice evolved from inclinations observable in our simian relatives" (1991, p. 335). In his view, such a sense of fairness (or justice[1]) entails not only the capacity of being aware of social rules but also of developing expectations about how certain interactions *should* take place. De Waal is confident that at least some non human animals possess both:

> I will describe behavior in chimpanzees and other primates that seems to reflect a sense of social regularity, that is, a sense of how others should or should not behave. This sense, which may be a precursor of the sense of justice, is defined here as a set of expectations about the way in which oneself (or others) should be treated and how resources should be divided, a deviation from which expectations to one's (or the other's) disadvantage evokes a negative reaction, most commonly present in subordinate individuals and punishment in dominant individuals. (1991, p. 336)

Bekoff and Pierce's more recent approach suggest a similar understanding of the phylogenetic roots of our own sense of fairness:

> A sense of justice is a continuous and evolved trait. And, as such, it has roots or correlates in closely related species or in species with similar

1. For merely practical reasons, in what follows I will refer to the issues related to 'justice' as issues of 'fairness'. When I speak of a 'sense of fairness', for example, I take is interchangeable with the expression 'sense of justice' used by De Waal, Bekoff, John Rawls and others.

> patterns of social organization. It is likely, of course, that a sense of justice is going to be species-specific and may vary depending on the unique and defining social characteristics of a given group of animals; evolutionary continuity does not equate to sameness. (2009, p. 115)

The implications involved in both approaches are important and far ranging: if it is truly the case that our sense of fairness is partly built on or rooted in certain dispositions or psychological mechanisms that we share with other species, then we cannot begin to understand the general structure and contents of our fairness evaluations until we have understood those shared elements. The presuppositions of this approach, however, are equally important and challenging, and it is those presuppositions that we need to delve into before we consider the possible connections between our fairness evaluations and other non human animals.

3. Sense of fairness: universal, innate, adapted, hypothetical

What do we mean exactly when we speak of the human *sense of fairness*? Do we mean that, as Nicolas Baumard (2016) suggests, we are endowed with a spontaneous tendency, akin to our senses of smell or touch, to distinguish between fair and unfair situations? Is it an 'innate instinct', as James Wilson (1997) suggests? A general definition that might perhaps be accepted by many of those who have written on the subject would be the following: the human sense of fairness is the spontaneous and unconscious tendency to evaluate situations in terms of fairness or unfairness. As is evident, however, since a mere definition is no proof of the actual existence of the thing defined, we can ask the following: do we really have such a tendency? Is there solid evidence to support the claim of its existence? If so, is it innate? If it is indeed innate and (therefore universal), is it modular in its architecture? If so, what type of modularity do we assume it possesses? More importantly: are we in conditions to address these questions given the present state of research concerning the genetic, physiological, and neurological bases of our fairness evaluations? Or are we still doomed to conform ourselves with one more just so story, built on evolutionary game theory approaches and/or theoretical considerations? I believe this last to be the case, since the important objections that have been raised concerning both the modularity thesis and the faculty-based approaches to neurology has led us to the brink of a Kuhnian crisis from which no new paradigm has yet emerged that can serve as a general framework from within which to articulate the investigations concerning the neurological bases of our sense of justice.

That being so, perhaps the best argument we can present for the moment for the existence of an innate sense of justice is, by way of a *reductio ad absurdum*, to challenge the reader to imagine a human being not endowed with the tendency to evaluate situations in terms of fairness: would it be plausible to assume that a human being could survive without any tendency whatsoever to defend himself against any unfair treatment or any tendency to resist being pushed around or taken advantage of? Barring the alternative of such a condition being a spandrel, the by-product of another trait that is itself adaptive[2], could natural selection favor the survival of such an individual? We can surely allow for such a possibility within the pages of a novel by Herman Melville, or within the tales of Jorge Luis Borges. We can even envisage it as one of the psychiatric study cases in Oliver Sacks' repertoire of singular characters, or as the (probably fleeting) result of a long and bizarre experiment in ascetic self-fashioning. But to believe that natural selection could leave the tendency to enforce fairness entirely in the hands of culture seems inconceivable in a cooperative and extremely social species as ours, since cooperation cannot evolve without the existence of the tendency to obey and enforce rules[3]. I believe that, at present, we are not able to produce a better account of the sense of fairness than that. There really isn't any evidence (as far as I am aware) to back up the claim that certain neural networks are responsible for the modular functioning of our tendency to evaluate situations in terms of fairness, and neither is there solid anthropological and historical evidence to support the existence of a universal sense of fairness. What we are left with is, in sum, the mere possibility of exploring theoretically an abstract proposal that may, perhaps, one day be definitely disproven or provisionally confirmed.

2. It is certainly impossible to rule out the emergence of our sense of fairness as a by-product of the emergence of consciousness, and that explanation would render the investigations in ethology mostly useless in our understanding of our fairness evaluations. What I propose to explore, however, is the (at least for now equally valid) opposite alternative.

3. It could be pointed out that in this particular instance, I am taking natural selection as operating on the individual level, but that it can also operate on a population level: alarm calls in non human primates, for example, entail a risk for the individual who utters them, but allows the rest of the population to escape a certain danger. The objection, however, would be irrelevant in this case on account of a structural difference between both scenarios: in the alarm call scenario, it is the *side effect of the possession of a certain trait* that entails a danger to its possessor; in the scenario I am discussing, the danger would come from *the absence of a certain trait* (the sense of fairness), and natural selection, as far as I know, does not select for absences.

4. Sense of fairness, fairness criteria and fairness evaluations

If our sense of fairness consists in the tendency to evaluate certain situations as fair or unfair, it is evident that our fairness evaluations presuppose the capacity to entertain concepts in at least two distinct but complementary instances: the evaluation itself of a certain situation *as fair or unfair*, and the criteria used in that evaluation (equality, reciprocity, property, etc.). In both cases we are attributing a certain property to a state of affairs and, in order to do so, we need the corresponding concepts. While the base evaluation can be considered universal (since there doesn't seem to be evidence of any culture that lacks a concept to denote general violations of norms), fairness criteria are partly cultural and optional: they are partly cultural in that certain societies or cultures may resort to specific criteria that others do not, and they are optional in that one may limit oneself to judge a certain situation as *unfair* without stating (or knowing) the criteria on which one has based one's judgment. In either case, the base evaluation does entail the use of the concepts of 'fair' and 'unfair', and can therefore only take place (as far as we know) in the mind of a human being: there may appear to be strong parallels or analogies between human and non human animals when it comes to enforcing norms, inflicting punishment, reciprocating favors, etc., but if fairness evaluations imply the concept of fairness, those parallels and analogies can only be superficial[4]. But if fairness evaluations are exclusively human, what light could the study of the behavior of non human animals possibly shed on our sense of fairness? If we adopt a thoroughly rupturist approach (as, for instance, Ayala (2010), Baumard (2016), and Korsgaard, (2006) do) the answer seems to be none: ethology has no insights to offer

4. It could be argued that although non human animals lack linguistic concepts, there are other approaches to the notion of "concept" that could be relevant in this respect, such as dispositionalists approaches (a recent review of the corresponding literature can be found in Danón (2021)). That would be not be the case here, however, since I have specifically defined our sense of fairness as "the spontaneous and unconscious tendency to evaluate situations *in terms of fairness or unfairness*", which explicitly entails the classification of a situation *as* "fair" or "unfair". If we were to define our sense of fairness as the spontaneous and unconscious tendency to evaluate situations *in ways that we humans could equate with our classifications of fair/unfair*, the objection would certainly be relevant. But that is not what I have done (mainly because it would involve determining which is the behavioral evidence in each case that would point to those non linguistic evaluations, an endeavor that I do not feel capable of accomplishing).

to those who are interested in studying our sense of fairness or human morality in general.

I believe that such a conclusion is not warranted if we distinguish clearly between the following elements: i) our fairness evaluations, ii) the psychological dispositions that underlie them, and iii) the criteria we (may) use to justify those evaluations[5]. These three elements, which are sometimes confounded in other approaches (leading to either anthropomorphism o anthropodenial) are clearly distinct in kind: our fairness evaluations are the actual and conscious evaluations of certain situations as fair or unfair, and are conceptual in kind, as are the criteria we may resort to when justifying those evaluations ("His decision was unfair because I had not done anything wrong!"). The psychological dispositions that underlie those evaluations, however, do not necessarily imply the use of concepts[6].

To be sure, fairness evaluations can be the result of careful and systematic deliberation: a judge may come to the conclusion that, taking into consideration certain criteria, a particular situation was unfair and therefore someone deserves to be punished. We can also come to similar conclusions after much reflection on something that has happened to us or someone else. In such cases, fairness evaluations do entail the conscious use of fairness criteria, and the study of the behaviors of chimpanzees, bonobos or dolphins appears to be completely irrelevant to further our understanding of the mental processes involved in those evaluations, in which logic (or culture) seems to have the final word. And the same seems to be the case when we deal with fairness evaluations that appear as the result of the unconscious application of fairness criteria: if, for example, I come to the conclusion that someone has evaluated a certain situation as unfair because the idea of failed reciprocity must have come into his

5. None of these elements must be confused with the *sense of fairness* itself, which is merely the (alleged) innate tendency to produce (i) *fairness evaluations* that may (or may not) be induced by the activation of certain (ii) *psychological dispositions*, and that may (or may not) be followed by a process of rationalization that resorts to certain (iii) *fairness criteria*. Even if it could be demonstrated that there is no such thing as an evolved and innate sense of fairness as I have defined it, the distinction between the remaining three elements would still hold.

6. One need not understand those innate dispositions as rigid 'action programmes' (as Damasio and Carvalho (2013) do); they can rather be thought of as some the building blocks that take part (alongside others) in certain behavioral responses. The fact that those dispositions interact with other (cultural or evolved) elements would then explain why individuals of the same species can react differently to a similar scenario.

mind, even if he hasn't consciously noted it, that evaluation is still (at least putatively) conceptual in nature. Fairness criteria clearly seem to come into play during unconscious processes and determine the outcome of mental processes that lead to fairness evaluations.

If we take into account the investigations that have been conducted from within dual process theories (Evans, 2008; Evans & Stanovich, 2013; Kahneman, 2012), however, fairness evaluations that are the actual result of the application of fairness criteria seem to be only half of the story, given that those evaluations can also be the result of mental processes that are (as far as we know) essentially affective in nature, and not conceptual[7]. If we further take into consideration the insights of Social Intuitionism (Haidt, 2001) or the interactionist approach to reason (Mercier & Sperber, 2017), conceptual evaluations appear to account for even less of the story, since the justifications we tend to give for our fairness evaluations seem to have less to do with the truth of what has led us to those evaluations and more with speculation or even fabulation. According a long line of research that dates back to Robert Zajonc and John Bargh (both influenced by William James' early insights), a fundamental consequence of the opacity of the inferential processes that take place in our mind is that we can only see the end result of those processes, but we cannot either witness nor reconstruct the actual process that has led to a certain conclusion. According to the Social Intuitionist Model proposed by Haidt (2001), when faced with such intuitions (i.e., conclusions that are the result of inaccessible processes[8]) we have a tendency not to simply admit our ignorance of why we feel in a certain way about something, but rather to produce *post hoc* rationalizations (without obviously not being aware that we are rationalizing – for if we were aware that we are doing so we would be consciously lying). Hugo Mercier & Dan Sperber (2017) further suggest that those rationalizations are based not on what we hold to be the *actual* causes of our evaluations, but rather the most *plausible* ones. A coincidence between the reasons that we think or declare that have led us to condemn a certain practice as unfair and the reasons (if any) that have actually led us to that conclusion, can only be the result of chance. Our declared reasons, in sum, are only probable guesses we

7. Although the distinction between rationality and emotions (or concepts and affects) has been questioned since at least the Hellenistic period, I believe it is still useful to pinpoint the psychological dispositions that take place in the mind of non linguistic species.

8. "Intuition is the best word to describe the dozens or hundreds of rapid, effortless moral judgments and decisions that we all make every day. Only a few of these intuitions come to us embedded in full-blown emotions" (Haidt, 2012, p. 60).

produce in order to fill the void on which our sense of fairness operates, a void that seems to be filled not only by culture, but also by evolution.

5. Cultural and evolved psychological dispositions

As working hypotheses, I have thus far suggested that is possible that:

i) we are born with what we may call a 'sense of fairness', which is a tendency to evaluate (in a spontaneous and unconscious manner) certain situations as fair or unfair;
ii) that our sense of fairness may be an adaptation;
iii) that although it may be modular (to some degree) we do not seem to be in a position at present to provide evidence concerning its physiological or neural bases.

I have also claimed that:

iv) our fairness evaluations can be the result either of the conscious[9] or unconscious application of certain criteria of fairness (which are conceptual in nature), or of certain intuitions whose inner workings elude us[10].

Is there room to believe that that fuzzy or dark space where our intuitions take place is inhabited by something other than merely cultural elements? Can natural selection have favored the presence of specific

9. Although there might seem to be a conflict between the proposed definition of our sense of fairness as unconscious and the admission that it may take as input the conclusions of a conscious deliberation, there is no such contradiction: the hypothesized tendency I label as sense of fairness is limited to the unconscious tendency to evaluate..., not to the actual evaluation. By way of analogy, we can think of the sensation of hunger that can sometimes lead to eat and the act of eating a specific food. Perhaps it would not be extremely inaccurate to push the analogy and think of our sense of fairness as an 'urge'.

10. After reviewing the relevant literature in the fields of psychology and philosophy, Hodgkinson et al. (2008) suggests the following as "a broadly consensual definition of intuition": "a complex set of inter-related cognitive, affective and somatic processes, in which there is no apparent intrusion of deliberate, rational thought. Moreover, the outcome of this process (an intuition) can be difficult to articulate. The outcomes of intuition can be experienced as an holistic 'hunch' or 'gut feel', a sense of calling or overpowering certainty, and an awareness of a knowledge that is on the threshold of conscious perception" (2008, p. 4). A clear and precise description of the multiple layers involved in the concept of intuition can be found in Scotto (2022).

psychological dispositions that explain certain tendencies to process information and react to it? Drawing in both cases from the tradition left behind by evolutionary psychology and by modular approaches to psychology, Haidt and his team think that that is certainly the case: nature seems to have endowed us with tendencies to react in specific manners to a set of inputs that are related to the adaptive challenges that our ancestors (both from and prior to the *homo sapiens* lineage) had to face in the past, and that explains the alleged regularities that we find below the surface of cultural diversity in terms of social rules and practices. Under this hypothesis, however, the innate and evolved dispositions that natural selection has provided us with goes far beyond the domain of fairness: we are probably born, for example, with a disposition tailored to recognize certain visual patterns (faces) and not others as evidence of the presence of a conspecific, but that disposition does not seem to have much to do with what we think of as fairness evaluations. The question, then, becomes the following: can our fairness evaluations draw on *any* evolved disposition? Or is it the case that only some of them serve as inputs to our sense of fairness? *Prima facie*, the second alternative seems to be the most plausible one: supposing, for example, that we are born with a general disposition to experience fear when faced with what appears to be a threat to our survival or well being, it would be strange to draw on that disposition and declare a situation that has produced fear in us as unfair. If it is true that we have an innate tendency to experience negative emotions towards what threatens the survival or well being of our offspring, it would be equally odd to consider such threats as unfair. I might try to do all I can to neutralize both threats and even condemn them *on other grounds*, but to label them as unfair would probably not be the most natural option.

But if it is true that our fairness evaluations do not draw on *any* evolved disposition, what determines which of the available dispositions are taken as inputs by our sense of fairness and which not[11]?

11. This, it must be stressed, is a descriptive question --- not a normative one: what we are asking is not whether the criteria or intuitions (cultural or evolved) that our sense of fairness relies on constitute a coherent set when taken together; nor if some of those criteria or intuitions are compatible (and others aren't) with a given abstract conception of justice that we consider on philosophical grounds to be the best available one. What we are asking, in other words, is if there is an objective (plausible) explanation to why our sense of fairness tends to rely on a certain set of dispositions (and not others) or to resort to certain criteria (and not other) – not whether it should do so from a logical, sociological or metaphysical point of view. I believe that it is precisely the failure to distinguish between both approaches that has prevented some researchers to comprehend the relevance of studying the relation between evolved dispositions and our sense of fairness.

6. Cooperation, reciprocity and fairness evaluations

> "She could never forget his kindness —
> he had been really remarkably kind —
> she forgot precisely upon what occasion.
> But he had been remarkably kind"
> (Virginia Woolf, *Mrs. Dalloway*)

Can the evolutionary origins of our species shed light on the previous question? Yes and no.

On the one hand, a look at our evolutionary origins certainly seems to explain the prominence and regularity that we witness in human history concerning, for example, the dynamics of reciprocity. If we assume that the cooperative nature of our species is the result of a process by which natural selection favored those individuals who showed certain dispositions that made (sustained and systematic) cooperation possible, the most basic dispositions of that sort seem to be, on the one hand, the tendency to reciprocate favors and, on the other, the tendency to adopt a negative attitude to those individuals who do not reciprocate. This does not imply, by any means, that cooperation presupposes the capacity to conceptualize rules as such: the tendencies favored by natural selection concerning cooperation may be as simple as a merely affective state that predisposes us favorably or unfavorably towards another, without any awareness of the nature and causes of that affective state. To be sure, on top of that basic initial negative attitude towards non reciprocation more complex phenomena can be built: one can feel a desire to behave aggressively towards a non reciprocator (but without actually doing so); one can effectively behave in such a manner; one can behave in such a manner while being aware of the aggressive nature of the reaction; and, at the end of the spectrum, one can be aware not only of the aggression but also of the fact that the aggression was a response to lack of reciprocation. This makes it evident that, on the one hand, complex phenomena, such as a conscious and explicit punishment of someone who has not reciprocated a favor, can be built on top of very simple evolved dispositions or tendencies, and, on the other, that those basic tendencies cannot be mistaken for their complex and cognitively demanding counterparts – a confusion that lies at the heart of the attribution of a (proto) sense of fairness to non human animals.

Jeffrey Stevens and Marc Hauser (2004) were perhaps among the first to explicitly address the question of the cognitive requirements of the mental traits that primatologists tended to attribute to non human animals. Concerning the specific case of 'reciprocal altruism' famously studied by Robert Trivers (1971), the authors claimed that several mental capacities had to be granted in a species (other than a cheater detection mechanism) for reciprocal altruism to take place: numerical quantification, temporal discounting, delayed gratification, analysis and recall of reputation, and inhibitory control. Although the authors' conclusion that reciprocal altruism must therefore be "rare if not absent among animals" (Stevens & Hauser, 2004, p. 64) seemed to condemn the investigations concerning reciprocity in animals to the museums of anthropocentrism, a more fruitful approach had already been sketched by De Waal several years before (but had remained largely ignored): in 2000, the author had put forward the notion of "attitudinal reciprocity" as a more parsimonious conceptual alternative to that of "calculated reciprocity", given that the former "is less cognitively demanding [...], because it does not assume mental score-keeping of given and received services nor expectations about appropriate return-favors, or the punishment of cheating" (De Waal, 2000, p. 260). De Waal's suggestion was later (implicitly or explicitly) taken over by several researchers: in response to Stevens and Hauser, Brosnan *et al.*, for example, resorted to the concept of attitudinal reciprocity as a phenomenon "in which individuals' responses are based on the positive feelings generated when a partner gives a favor, not on an exact accounting of favors given and received" (2009, p. 595). A similar approach was taken by Gomes et al. (2009), who put forward the idea that attitudinal reciprocity might involve the release of oxytocin, and by Gabriele Schino and Filippo Aureli, who suggested that De Waal's mechanism need not be limited to immediate interactions, but could be the basis of "a system of emotionally based bookkeeping that allows the long-term tracking of reciprocal exchanges with multiple partners without causing an excessive cognitive load" (Schino & Aureli, 2009, p. 5).

The substitution of the mechanism of 'calculated reciprocity' for that of 'attitudinal reciprocity' represents a paradigmatic example of the adoption of a non anthropomorphic approach to the mental lives of non human animals, in that it not also manages to put as little cognitive burden as possible on the subjects involved, but also has the additional advantage of not demanding, on the level of proximate explanations, anything more

than an affective state, rather than planification and forethought[12]. This last point is crucial, since it allows us to understand how reciprocity patterns can arise without any awareness of the idea of reciprocity, or, in other terms, how it is possible to find regularities without rules[13]. And it also explains why the dispositions that make reciprocity patterns stable can well predate our species, making research in cognitive ethology relevant to further our understanding of our fairness evaluations.

Taking into consideration how the emergence of cooperation may have favored the emergence of certain dispositions that feed into our sense of fairness, however, has important limits in explaining which evolved dispositions our sense of fairness takes as input and which ones it doesn't. And the reason for that is that reciprocity represents only one of the several criteria that humans resort to when explicitly arguing in

12. As the long debate concerning Brosnan & De Waal's famous paper 2003 shows, "minimalistic" approaches carry more plausibility when suggesting interpretations of complex mental phenomena: in this particular case, it has been suggested that instead of assuming that Capuchin monkeys have the cognitive capacity to "measure reward in relative terms" (p. 299), it is enough to resort to the "frustration effect", which (following Abram Amsel's (1958) seminal paper on the subject) has been studied extensively in several non human species (Dantzer et al., 1980; Freidin & Mustaca, 2004; Jakovcevic et al., 2013; McPeake et al., 2021; Papini, 2003; Papini et al., 2019; Stout et al., 2003; Wilton et al., 1969). Although the pattern of responses elicited by the frustration of expectations varies across species, certain observed regularities, such as the increase in aggression or agonistic behavior, seem to account more parsimoniously for at least some of the behaviors in non human animals that might at first sight be interpreted as (proto) fairness evaluations (Brauer et al., 2006; Cheney & Seyfarth, 2007; Dubreuil et al., 2006; Fletcher, 2008; Roma et al., 2006; Silberberg et al., 2009). A similar approach can be taken when proposing interpretations concerning the proximate causes of behaviors that appear on the surface to express an intention to "punish" norm violations, as has been suggested by Raihani et al. (2012) and Riedl et al., (2012).

13. De Waal's first explicit attribution of a sense of fairness to non human animals relied on this confusion: "All animals conform to social rules. That is, their conduct toward their con-specifics is to some degree predictable." (1991, p. 337). That the behavior of an individual is predictable does not entail that he is conforming to a certain rule. Charlotte Hemelrijk, for example, had already suggested that certain regularities in reciprocation that had been interpreted as evidence of Reciprocal Altruism in chimpanzees, could be explained as a side reciprocity and interchange may arise "as a side-effect from self-reinforcing aggressive interactions, spatial structure and grooming between artificial entities that lack every motivation to reciprocate" (1997, p. 190). Concerning the transactions that follow a collective hunt by a group of chimpanzees, which had previously been approached, for example, through the "meat for sex" hypothesis, Ian Gilby proposed that such transactions could be explained by resorting to the variable of harassment: the more an individual insistently claims for a share of the booty, the more chances of success he will have (2006, p. 20). From this perspective, the decision of the chimpanzee that has led the hunt to give up part of the booty does not obey a principle of reciprocity or fairness, but can be explained by factors that have nothing to do with a sense of fairness.

favor of the fairness or unfairness of a certain outcome. As Fiske and Rai (2015) have shown, for example, our fairness evaluations tend to take as input an extraordinary range of dispositions and cultural elements, which can go from violations of authority ranking to alleged divine dispositions concerning the right of some to possess more than others.

Does this mean, once again, that our fairness evaluations can be triggered by any disposition whatsoever? One of the most recent and systematic attempts to answer this question has been Haidt' Moral Foundation Theory, which claims that the whole range of dispositions[14] we are endowed with can be probably classified under six moral domains (care, fairness, loyalty, authority, sanctity, liberty) that are clearly distinguished by the specific evolutionary problems they were supposed to answer: in the case of the foundation of care/harm, the problem to be solved is that of protecting and caring for our offspring, in the case of the loyalty foundation, to form cohesive coalitions, and so on (Haidt, 2012, p. 139).

Haidt's theory (which explicitly draws inspiration from Fiske's Relational Models Theory) is certainly helpful in sorting out the possible evolutionary challenges faced by our ancestors and the possible dispositions that may have contributed to solving those challenges, but it faces two important difficulties: on the one hand, the criteria that Haidt and his team have proposed in order to establish the limits of each domain are not particularly clear, since at least some of the different evolutionary challenges proposed can be linked to more than one of the dispositions analyzed by them. On the other hand, the relation proposed by their model between dispositions and foundations is, as I have argued elsewhere (Braicovich, 2021), particularly vague, since the theory still fails to answer why our sense of fairness (or any of the other domains) tends to take as input certain intuitions and not others. The failure to answer that question, I believe, is, on the one hand, due to the fact that Haidt has explicitly borrowed from Sperber's distinction between the proper and the actual domain of a module, which states that a module can come to be triggered by a set of conditions that are different from the original ones under which it was shaped and proved adaptive. (So much so, in fact, that "the actual domain of any human cognitive module is unlikely to be even approximately coextensive with its proper domain from versions of modularity" (Sperber, 1994, p. 54)).

14. Although Haidt refers to them as 'emotions', I will keep the term disposition when referring to Moral Foundation Theory in order to preserve the points of contact between my proposal and his. After all, Haidt does acknowledge that the cognitive modules he is referring to have a considerable range of expressions, from mere 'flashes', to 'affective reactions' and proper 'emotions', the last of which apply to instances when a certain foundation is "activated strongly" (Haidt, 2012, p. 140).

Both Haidt's distinction between the original and the current triggers of a domain and Sperber's distinction between the proper and the actual domains of a module, in sum, are extremely unclear concerning the relation between the original/proper and the current/actual domains, which might lead one to believe that almost anything can come to trigger, for example, our sense of fairness. But that is precisely what seems to be the case when we look at the problem from a historical and ethnographical perspective, and it is that essential indefinition what makes both approaches relevant and interesting. And that shows, incidentally, why a look at the evolutionary history of our species (particularly concerning the emergence of cooperation) can only shed a partial light on the mental processes that underlie our fairness evaluations, leaving the rest of the work to sociology, ethnography and cultural studies[15].

7. Final remarks

On the question of whether what we term 'sense of fairness' is an exclusively human feature or whether it can be traced to other non human species with which we share a common ancestry, I suggested the following (highly speculative) argument, which is composed of two main independent premises (1 and 4):

1. Random genetic mutations in certain individuals of non human species determined the presence in those individuals of certain innate dispositions that proved favorable to cooperative strategies.
2. As descendants of those species, we inherited those innate dispositions because of the process of natural selection.
3. Those evolved (innate) dispositions tend to trigger (in response to different stimuli) variations in our affective state (either positive or negative) which we are not (necessarily) aware of.
4. (Perhaps) for evolutionary reasons, we, as a species, are born with a tendency to evaluate intersubjective scenarios (mainly but not exclusively) in terms of 'fairness' and 'unfairness'.

15. While I doubt whether ethological studies can shed light on our sense of fairness (given that it entails the use of concepts and can be therefore deemed as an exclusively human trait – if it exists at all), I believe that cognitive ethology can certainly help us understand the evolved psychological dispositions that we have inherited form our ancestors and that still operate as (unconscious) inputs of many of our fairness evaluations. To that extent, although the contribution that cognitive ethology can make to our understanding of our fairness evaluations is only partial, it is a definite and precise contribution: we cannot possibly begin to understand those evaluations until we admit that they are too often built (among other things) on top of psychological dispositions that are shared with non human animals.

5. Those fairness evaluations can be the result of (at least) two types of mental processes: conscious deliberation and intuitions, the latter being of two types: cultural or evolved.
6. Our sense of fairness, in other words, can take as input either cultural elements or the variations in our affective state that are the result of a set of innate dispositions that were selected for due to their adaptive value concerning cooperation.

As is (controversially) evident, the main premises of this argument rest on the assumption that some version of innatism is still defensible, and that natural selection operated (at least in our evolutionary lineage) at the level of individual dispositions, rather than favoring general-purpose cognitive mechanisms, which applies not only to our 'sense of fairness', but also to the set of evolved dispositions that lie at the basis of, for example, attitudinal reciprocity or the frustration effect.

Although those dispositions can become resignified and integrated into fairness evaluation, they do not imply in themselves any notion of fairness. It certainly seems apt to think of them as the "building blocks" of the tower of morality, as De Waal has suggested, but bearing in mind that the morality (or fairness) part of the divided line starts with concepts. As the examples of attitudinal reciprocity and the frustration effect show, it is possible that the result of those dispositions is merely a rather undefined affective state with positive or negative valence and with varying intensity – affective states that do not imply either awareness or the use of concepts, but can nevertheless easily merge with other co-occurring processes that are themselves conceptual in nature. From this perspective, the effect of the frustration of our expectations, or the non reciprocation of a favor, to cite only two examples, can combine with a wide array of cultural elements (such as the multiple fairness criteria studied by Alan Page Fiske and Tage Shakti Rai (2015)) to deliver fairness evaluations that may seem at first sight to have little or nothing to do with non conceptual processes. And it is the process of rationalization that serves to obscure the indelible stamp of the lowly origins of our most sophisticated and elaborate fairness evaluations[16]: as Robert Solomon puts it, "much of what goes under the name of 'justice' is, one could argue, an elaboration of certain familiar feelings not dissimilar to vengeance (other, obviously, than a polar shift in 'valence') rather than a grand scheme or blueprint for the rational organization of society" (1995, p. 255).

16. We can advance a further question here: does rationalization possess an adaptive value? Mercier, Sperber and Baumard certainly seem to think so. See Baumard (2016), Mercier & Sperber (2017), and Sperber & Baumard (2012).

Understanding the central role that rationalization plays when giving account of the rationale behind our fairness evaluations, allows us to sever the causal tie between what we declare to be the motives behind those evaluations and its actual causes. Although the explicit justification of our evaluations may (by mere chance) coincide with their actual processes, we are in no way of actually knowing if they are or not, neither as philosophers, neuroscientists or psychologists, nor as regular human beings, since folk psychology, introspection and intuitive metaphysics do not seem to be better suited to explain the mental processes that our theoretical models have failed so far to account for.

But the main advantage of integrating (cultural and evolved) intuitions and rationalization into our understanding of fairness evaluations is that it allows us to approach them from a perspective that grants a role as important to culture as to evolution, preserving the exclusivity of our sense of fairness while, at the same time, acknowledging that it can frequently operate on top of evolved mechanisms that predate our species.

References

Amsel, A. (1958). The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, *55*(2), 102-119. https://doi.org/10.1037/h0043125

Ayala, F. J. (2010). The difference of being human: Morality. *Proc. Nati. Acad. Sci. USA*, *107*(Supl. 2), 9015-9022. https://doi.org/10.1073/pnas.0914616107

Baumard, N. (2016). *The Origins of Fairness: How Evolution Explains our Moral Nature*. New York: Oxford University Press.

Bekoff, M., & Pierce, J. (2009). *Wild justice: The moral lives of animals*. Chicago: The University of Chicago Press. https://doi.org/10.7208/chicago/9780226041667.001.0001

Braicovich, R.S. (2021). La Antropología Filosófica frente al factum de la evolución. In R. López-Orellana, J. & J. Suárez-Ruíz (Eds.), *Filosofía posdarwiniana. Enfoques actuales sobre la intersección entre análisis epistemológico y naturalismo filosófico* (pp. 337-355). College Publications.

Brauer, J., Call, J., & Tomasello, M. (2006). Are apes really inequity averse? *Proceedings of the Royal Society B*, *273*, 3123-3128. https://doi.org/10.1098/rspb.2006.3693

Brooks, J. A., Shablack, H., Gendron, M., Satpute, A. B., Parrish, M. H., & Lindquist, K. A. (2017). The role of language in the experience and perception of emotion: A neuroimaging meta-analysis. *Social Cognitive*

and Affective Neuroscience, *12*(2), 169-183. https://doi.org/10.1093/scan/nsw121

Brosnan, S. F., & De Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature*, *425*, 297-299. https://doi.org/10.1038/nature01963

Brosnan, S. F., Silk, J. B., Henrich, J., Mareno, M. C., Lambeth, S. P., & Schapiro, S. J. (2009). Chimpanzees (Pan troglodytes) do not develop contingent reciprocity in an experimental task. *Animal Cognition*, *12*(4), 587-597. https://doi.org/10.1007/s10071-009-0218-z

Buller, D. J., & Hardcastle, V. G. (2000). Evolutionary Psychology, meet Developmental Neurobiology: Against promiscuous modularity. *Brain and Mind*, *1*, 307-325. https://doi.org/10.1023/A:1011573226794

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon metaphysics: The evolution of a social mind*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226102429.001.0001

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: Evolutionary and neurobiological origins. *Nature Reviews Neuroscience, 14*(2), 143-152. https://doi.org/10.1038/nrn3403

Danón, L. (2021). Conceptos en animales no humanos. In *Enciclopedia de la Sociedad Española de Filosofía Analítica*. http://www.sefaweb.es/conceptos-en-animales-no-humanos

Dantzer, R., Arnone, M., & Mormede, P. (1980). Effects of frustration on behaviour and plasma corticosteroid levels in pigs. *Physiology & Behavior*, *24*(1), 1-4. https://doi.org/10.1016/0031-9384(80)90005-0

De Waal, F. B. M. (1991). The chimpanzee's sense of social regularity and its relation to the human sense of justice. *American Behavioral Scientist*, *34*(3), 335-349. https://doi.org/10.1177/0002764291034003005

De Waal, F. B. M. (2000). Attitudinal reciprocity in food sharing among brown capuchin monkeys. *Animal Behaviour*, *60*(2), 253-261. https://doi.org/10.1006/anbe.2000.1471

Dubreuil, D., Gentile, M., & Visalberghi, E. (2006). Are Capuchin Monkeys (Cebus Apella) Inequity Averse? *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1223-1228. https://doi.org/10.1098/rspb.2005.3433

Dupré, J. (2001). *Human nature and the limits of science*. Oxford: Oxford University Press. https://doi.org/10.1093/0199248060.001.0001

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Feldman Barrett, L. (2017). *How emotions are made: The secret life of the brain*. Boston: Houghton Mifflin Harcourt.

Feldman Barrett, L., & Russell, J. A. (Eds.). (2015). *The psychological construction of emotion*. New York: The Guilford Press.

Fiske, A. P., & Rai, T. S. (2015). *Virtuous Violence: Hurting and Killing to Create, Sustain, End, and Honor Social Relationships*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316104668

Fletcher, G. E. (2008). Attending to the outcome of others: Disadvantageous inequity aversion in male capuchin monkeys (Cebus apella). *American Journal of Primatology*, *70*(9), 901-905. https://doi.org/10.1002/ajp.20576

Fox, P. T., & Friston, K. J. (2012). Distributed processing; distributed functions? *NeuroImage*, *61*(2), 407-426. https://doi.org/10.1016/j.neuroimage.2011.12.051

Freidin, E., & Mustaca, A. E. (2004). Frustration and sexual behavior in male rats. *Animal Learning & Behavior*, *32*(3), 311-320. https://doi.org/10.3758/BF03196030

George, N., & Sunny, M. M. (2019). Challenges to the Modularity Thesis Under the Bayesian Brain Models. *Frontiers in Human Neuroscience*, *13*, 353. https://doi.org/10.3389/fnhum.2019.00353

Gilby, I. C. (2006). Meat sharing among the Gombe chimpanzees: Harassment and reciprocal exchange. *Animal Behaviour*, *71*(4), 953-963. https://doi.org/10.1016/j.anbehav.2005.09.009

Gomes, C. M., Mundry, R., & Boesch, C. (2009). Long-term reciprocation of grooming in wild West African chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1657), 699-706. https://doi.org/10.1098/rspb.2008.1324

Gould, S. J. (1996). *The mismeasure of man* (Rev. and expanded). New York: Norton.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834. https://doi.org/10.1037//0033-295X.108.4.814

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Vintage Books.

Hemelrijk, C. K. (1997). Reciprocation in apes: From complex cognition to self-structuring. En W. C. McGrew, L. F. Marchant, & T. Nishida (Eds.), *Great Ape Societies* (pp. 185-195). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511752414.016

Hodgkinson, G. P., Langan-Fox, J., & Sadler-Smith, E. (2008). Intuition: A fundamental bridging construct in the behavioural sciences. *British Journal of Psychology*, *99*, 1-27. https://doi.org/10.1348/000712607X216666

Jakovcevic, A., Elgier, A. M., Mustaca, A. E., & Bentosela, M. (2013). Frustration behaviors in domestic dogs. *Journal of Applied Animal Welfare Science*, *16*(1), 19-34. https://doi.org/10.1080/10888705.2013.740974

Kahneman, D. (2012). *Pensar rápido, pensar despacio*. Buenos Aires: Debate.

Korsgaard, C. M. (2006). Morality and the Distinctiveness of Human Action. En S. Macedo & J. Ober (Eds.), *Primates and Philosophers: How Morality Evolved* (pp. 98-119). Princeton: Princeton University Press. https://doi.org/10.1515/9781400830336-008

Levins, R., & Lewontin, R. C. (2009). *The dialectical biologist*. Delhi: Aaker Books.

McPeake, K. J., Collins, L. M., Zulch, H., & Mills, D. S. (2021). Behavioural and Physiological Correlates of the Canine Frustration Questionnaire. *Animals*, *11*(12), 3346. https://doi.org/10.3390/ani11123346

Mercier, H., & Sperber, D. (2017). *The Enigma of Reason: A New Theory of Human Understanding*. Cambridge: Harvard University Press. https://doi.org/10.4159/9780674977860

Mosterín, J. (2011). *La Naturaleza humana*. Madrid: Espasa.

Palecek, M. (2017). Modularity of Mind: Is It Time to Abandon This Ship? *Philosophy of the Social Sciences*, *47*(2), 132-144. https://doi.org/10/gm5rzb

Papini, M. R. (2003). Comparative Psychology of Surprising Nonreward. *Brain, Behavior and Evolution*, *62*(2), 83-95. https://doi.org/10.1159/000072439

Papini, M. R., Penagos-Corzo, J. C., & Pérez-Acosta, A. M. (2019). Avian Emotions: Comparative Perspectives on Fear and Frustration. *Frontiers in Psychology*, *9*, 1-14. https://doi.org/10.3389/fpsyg.2018.02707

Pinker, S. (2002). *The Blank Slate: The Modern Denial of Human Nature*. New York: Penguin.

Prinz, J. J. (2006). Is the mind really modular? En R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 22-36). Malden: Blackwell.

Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. *Trends in Ecology & Evolution*, *27*(5), 288-295. https://doi.org/10.1016/j.tree.2011.12.004

Riedl, K., Jensen, K., Call, J., & Tomasello, M. (2012). No third-party punishment in chimpanzees. *Proceedings of the National Academy of Sciences*, *109*(37), 14824-14829. https://doi.org/10.1073/pnas.1203179109

Roma, P. G., Silberberg, A., Ruggiero, A. M., & Suomi, S. J. (2006). Capuchin monkeys, inequity aversion, and the frustration effect. *Journal of Comparative Psychology*, *120*(1), 67-73. https://doi.org/10.1037/0735-7036.120.1.67

Sahlins, M. (2008). *The Western illusion of human nature*. Chicago: Prickly Paradigm Press.

Schino, G., & Aureli, F. (2009). Reciprocal altruism in primates. Partner choice, cognition, and emotions. *Advances in the Study of Behavior*, *39*, 45-69. https://doi.org/10.1016/S0065-3454(09)39002-6

Scotto, S. C. (2022). Cognición moral y cognición psicológica: Las intuiciones vienen primero. *Revista de Humanidades de Valparaíso*, *19*, 15-42. https://doi.org/10.22370/rhv2022iss19pp15-42

Silberberg, A., Crescimbene, L., Addessi, E., Anderson, J. R., & Visalberghi, E. (2009). Does inequity aversion depend on a frustration effect? A test with capuchin monkeys (Cebus apella). *Animal Cognition*, *12*(3), 505-509. https://doi.org/10.1007/s10071-009-0211-6

Solomon, R. C. (1995). Justice as vengeance, vengeance as justice. A partial defense of Polymarchus. En J. P. Sterba (Ed.), *Morality and Social Justice: Point/counterpoint*. Maryland: Rowman & Littlefield.

Sperber, D. (1994). The modularity of thought and the epidemiology of representations. En L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind* (pp. 39-67). Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9780511752902.003

Sperber, D., & Baumard, N. (2012). Moral Reputation: An Evolutionary and Cognitive Perspective: Moral Reputation. *Mind & Language*, *27*(5), 495-518. https://doi.org/10.1111/mila.12000

Stevens, J. R., & Hauser, M. D. (2004). Why be nice? Psychological constraints on the evolution of cooperation. *Trends in Cognitive Sciences*, *8*(2), 60-65. https://doi.org/10.1016/j.tics.2003.12.003

Stout, S. C., Boughner, R. L., & Papini, M. R. (2003). Reexamining the frustration effect in rats: Aftereffects of surprising reinforcement and nonreinforcement. *Learning and Motivation*, *34*(4), 437-456. https://doi.org/10.1016/S0023-9690(03)00038-9

Suhler, C. L., & Churchland, P. (2011). Can Innate, modular «foundations» explain morality? Challenges for Haidt's Moral Foundations Theory. *Journal of Cognitive Neuroscience*, *23*(9), 2103-2116; discussion 2117-2122. https://doi.org/10/fsr92d

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35-57. https://doi.org/10.1086/406755

Wilson, J. Q. (1997). *The moral sense*. New York: Free Press.

Wilton, R. N., Strongman, K. T., & Nerenberg, A. (1969). Some Effects of Frustration in a Free Responding Operant Situation. *Quarterly Journal of Experimental Psychology*, *21*(4), 367-380. https://doi.org/10.1080/14640746908400232