Revista de Humanidades de Valparaíso, 2022, No 19, 63-83 *Humanities Journal of Valparaíso*
DOI: https://doi.org/10.22370/rhv2022iss19pp63-83 eISSN 0719-4242
Sección Monográfica / Monographic Section

The Emotional Dog Was a Glauconian Canine: The Reception of the Social Intuitionist Model, From the Neurocentric Paradigm to the Digital Paradigm

El perro emocional era un cánido glauconiano: la recepción del modelo intuicionista social, del paradigma neurocéntrico al paradigma digital

Pedro Jesús Pérez Zafrilla
Universidad de Valencia, España
p.jesus.perez@uv.es

Abstract

In this article I analyze the academic reception of Jonathan Haidt's seminal article *The emotional dog and its rational tail: A social intuitionist approach to moral judgment*. My thesis is that in the spheres of philosophy and psychology, this article was initially studied within the neurocentric paradigm, which dominated the field of scientific reflection in the fifteen years following its publication. This neurocentric reading established a specific interpretation of the text with several limitations. However, more recently a digital paradigm has emerged and come to prevail in academia, providing a new perspective from which to return to Haidt's text. Indeed, this approach makes it possible to unravel elements of the famous article that in the neurocentric paradigm went unnoticed by researchers. Moreover, the digital paradigm manages to better integrate Haidt's seminal article into his later work as a whole.

Keywords: social intuitionist model, neurocentric paradigm, digital paradigm, social networks.

Resumen

En este artículo analizo la recepción que se ha hecho en la academia del artículo seminal de Haidt "The emotional dog and its rational tail: A social intuitionist approach to moral judgment". Mi tesis es que el estudio de este artículo en los campos de la filosofía y la psicología en un primer momento se llevó a cabo dentro del paradigma

Received: 31/04/2021. Final version: 26/03/2022
eISSN 0719-4242

neurocéntrico que, durante los siguientes quince años posteriores a la publicación del artículo, dominó el ámbito de la reflexión científica. Esta lectura neurocéntrica marcará una interpretación concreta del texto que presenta varias limitaciones. Sin embargo, el actual paradigma de reflexión sobre el mundo digital, imperante en el ámbito académico, conforma una nueva perspectiva desde la que volver al texto de Haidt. Este nuevo enfoque permite desentrañar elementos del famoso artículo que en el paradigma neurocéntrico pasaron desapercibidos para los investigadores. Además, el paradigma digital logra integrar mejor el artículo seminal de Haidt en el conjunto de su obra posterior.

Palabras clave: modelo intuicionista social, paradigma neurocéntrico, paradigma digital, redes sociales.

1. Introduction

Those of us whose philosophical research focuses on the fields of neuroethics and neuropolitics are well aware of the significance of the article *The emotional dog and its rational tail: A social intuitionist approach to moral judgment* (Haidt 2001) in shaping these disciplines' development. However, being twenty years since the publication of Jonathan Haidt's seminal article, we can also now take stock of how the interpretation of this text has evolved in academia.

In this article I argue that the reading of Haidt's article has been conditioned by two factors: by the very wording and argumentative structure that Haidt gave to the article, and by the neurocentric paradigm in which that interpretation was framed. It has been precisely with the change of framework in academia as a result of the emergence of the digital paradigm that, in my opinion, it is possible to return to the article and make a different reading of it.

First, I will place the social intuitionist model within the framework of neuroethical reflection. Then, I will present some hermeneutical keys to *The emotional dog and its rational tail* which, in my opinion, determined the article's reception by its critics in the neurocentric paradigm, and subsequently present the reading of Haidt's text in that paradigm. I will describe the main criticisms made of the article, both in the Anglo-Saxon and in the Spanish-speaking world, placing special emphasis on those made by the Applied Ethics and Democracy Research Group (also known as School of Valencia), to which I belong. Finally, I will discuss how in recent years the emergence of social networks as well as theories of evolutionary psychology has laid the foundations from which to articulate a new approach to *The emotional dog and its rational tail*. I call this innovative approach the "digital paradigm." This new approach makes it possible to unravel elements of the famous article that in the neurocentric paradigm went unnoticed by researchers. Moreover, the digital paradigm manages to better integrate Haidt's seminal article into his later work as a whole.

2. The social intuitionist model in the field of neuroethics

The publication of *The emotional dog and its rational tail* in 2001 coincided with the emergence of the neurocentric turn in academia, which owed to the development of neuroimaging techniques during the preceding decade. The possibility of knowing in real time how the brain works led various neuroscientists to explain the spheres of society (e.g., economics, politics, religion, ethics) based on the study of neuronal activation. Thus, neurodisciplines such as neuroeconomics, neuropolitics, neurotheology and neuroethics emerged in the 2000s (Cortina 2011).

Neuroethics, which is carried out by psychologists, neurolinguists, neuroscientists and philosophers, aims to create an *ethics of neuroscience*, an ethical framework to regulate neuroscientific research. However, it also includes a *neuroscience of ethics*: the study of neuroscientific discoveries' consequences for our understanding of ethics as well as people's behavior and moral agency (Roskies 2002). This second branch of research is the most substantial. It reflects on aspects such as the possibility of freedom and the affective and non-rational nature of morality, as well as on the evolutionary origin of moral judgment and reasoning. It is precisely in such analysis of the character of moral judgment that Haidt's work fits into the field of neuroethics.

Haidt's social intuitionist model is framed within what is known as the dual process model developed in social psychology. The dual process model emerged in the 1970s to explain moral cognition, and displaced the dominant rationalist model represented by Kohlberg and Turiel. The dual process model was developed throughout the 1990s, and is today the prevailing paradigm in social and cognitive psychology. Authors such as Nissbet and Wilson (1977), Margolis (1987), Zajonc (1980) and Kahneman (2011) defend the relevance of distinguishing two forms of moral cognition. Specifically, on the one hand, there is intuition, which is represented as rapid, unconscious, automatic and effortless. This mechanism is called System 1. On the other hand, there is reasoning, which is a slow, conscious, controlled and effortful process (called System 2). For all these authors, System 1 has primacy over System 2 in the formation of moral judgments.

It is within this framework, and on the basis of the studies of Nissbet and Wilson, Margolis, and Zajonc, that Haidt articulated his proposed social intuitionist model to account for moral judgment and reasoning. However, it was also within this context of neurocentric reflection on the implications of the dual process regarding the nature of moral cognition that *The emotional dog and its rational tail* was received in the fields of psychology and philosophy.

3. Hermeneutical keys to "The emotional dog and its rational tail"

In order to better understand the criticisms made of the 2001 article within the neurocentric paradigm, I will first focus on Haidt's presentation of his theses in the article. In my opinion, in the structure of the article we can find one of the hermeneutical keys for understanding critics' reading of the text. I argue that Haidt's article can be clearly divided into two parts:

- The first part is longer, running from the beginning up to the section "Four reasons to doubt the causal importance of reason." It aims to present the characteristic elements of the social intuitionist model in response to the rationalist model. It also includes the model's explanation of evidence against the causal effect of reasoning in moral judgment.
- The second part of the article goes from the section "The mechanism of intuition" to the end of the text. It seeks to explain the origin of intuitions and how they are cultivated in the social environment. It also tries to address the relationship between reasoning and intuition in the social intuitionist model.

The two parts clearly differ on four fundamental points:

- The first is methodological in nature. Specifically, whereas the first part of the text has an expository character of the main elements of the social intuitionist model as opposed to rationalism, the second part is more directed to the model's foundation, appealing to other theories already established in academia.
- The second difference affects the content. Thus, whereas the first part of the article deals, in an expository way, with the theory of judgment and moral reasoning within the social intuitionist model, the second part analyzes the adaptive origin of intuitions.
- The third difference concerns the discipline on which Haidt bases his arguments. Indeed, whereas the first part is centered on social psychology, as Haidt discusses moral judgment and moral reasoning in the social intuitionist model and does so in dialogue with the rationalist theory, in the second part Haidt enters into dialogue with other disciplines, such as anthropology and primatology.
- The fourth and final difference between the two parts of the article is related to the heuristic key that guides the argumentation in each of them. Specifically, the first part of the article responds to an argumentative logic that we can qualify as individualistic. That is to say, it constantly discusses how moral judgment and reasoning work at the individual level and in relationships between individuals. By contrast, the second part has a group logic. Here the focus is no longer on the psychological processes that take place in the minds of individuals; rather, the focus is on how morality develops as a cohesive phenomenon within a group.

To show the existing differentiation between these two parts of the article, I will refer to Haidt's characterizations in both of them of judgment, reasoning, morality and, above all, social character, which even gives its name to his neuropsychological theory. I will begin with the latter.

3.1. The social dimension

Haidt defines "social" as synonymous with "interpersonal": "the social part of the social intuitionist model proposes that moral judgment should be studied as an interpersonal process" (Haidt 2001, 814). Thus, the social dimension seems to be reduced to the interaction between subjects. This is also indicated in the figure illustrating the social intuitionist model, in which the social part is represented by link 4 (the social persuasion link, when the judgment of subject A influences the intuition of subject B) and link 3 (the reasoned persuasion link, when the reasoning of A influences the intuition of B). Thereby, "social" refers to interaction to influence another person's intuitions.

However, this characterization contrasts with other statements Haidt makes throughout the article, in which "social" is better understood as "group membership." For example, in referring to relatedness motives, Haidt notes that evolutionarily it would be disastrous if the machinery of moral judgment were designed to seek the truth rather than to agree with our friends over our enemies (Haidt 2001, 821). Furthermore, Haidt argues that "individuals must use language to influence others while simultaneously being at least somewhat open to interpersonal influence as specific norms, values, or judgments spread through a community" (Haidt 2001, 826).

The contrast between this last quotation and the above definition of "social" as "interpersonal" indicates the existence of an amphibology in the concept of "interpersonal." In the first text, "interpersonal" refers to the merely intersubjective, reducing the social to a relationship between subjects. Instead, in the second passage, "interpersonal" is understood as the place occupied by the individual in relation to a community. Thus "interpersonal" (and, therefore, social) comes to be interpreted as group membership. This is a discrepancy that has significant consequences for the interpretation of the social intuitionist model.

3.2. Intuition

In the first part of the text, intuition is analyzed from social psychology and is represented as one of two cognitive processes (together with reasoning) that make up the dual process model. Here Haidt defines intuition as "the sudden appearance in consciousness of a moral judgment, including an affective valence (good-bad, like-dislike), without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion" (Haidt 2001, 818). Thus, Haidt understands intuition according to the dual process model as an automatic and unconscious process, as opposed to reasoning, which is

effortful and conscious. Intuition is also a process that arises in the individual in response to the influence of other subjects, either through their reasoning or as a function of a series of biases.

This characterization of intuition as a psychological process is completed in the second part of the text with an exposition of how intuitions have an evolutionary origin and are formed and cultivated in a social context. That is to say, the second part explains how intuitions are not mere psychological reactions arising in the individual in response to the influence of other subjects; rather, they make sense within the framework of cultural practices and customs, and arise in response to problems derived from living together within groups.

Consequently, there is a contrast between two approaches. In the first part of Haidt's article, a purely individualistic and psychologistic approach prevails. Here intuition is a mere psychological process that occurs in the mind of the subject in response to an external stimulus. By contrast, the second part presents a group approach, in which intuition is a reaction that makes sense within a culture.

3.3. Moral judgment

In his exposition of the social intuitionist model, Haidt states that moral judgments consist of an evaluation of liking or disliking (i.e., good or bad) which appears in the mind, but of whose formation the subject is not conscious, because such judgments are the result of an intuition. Thus, "[t]he model proposes that moral judgments appear in consciousness automatically and effortlessly as the result of moral intuitions" (Haidt 2001, 818). Haidt goes so far as to affirm that this intuitive origin of moral judgments (and the *post hoc* character of reasoning) is the central thesis of the social intuitionist model (Haidt 2001, 817), in clear opposition to the rationalist theory. Moreover, the entirety of the first part of the article is intended to defend how most moral judgments have an intuitive rather than a reflexive origin. That is to say, a moral judgment is, in most situations, an affective evaluation caused in the subject by an intuition resulting from an influence external to the subject.[1]

However, the very definition of moral judgment, which is provided even in the first part of the article, does not reduce judgment to an evaluation of the approval or rejection of something caused by an external influence. Rather, moral judgments are defined as "evaluations (good vs. bad) of the actions or character of a person that are made with respect to a set of virtues held to be obligatory by a culture or subculture" (Haidt 2001, 817). According to this definition, a moral judgment cannot be reduced to the appearance in consciousness of a mere intuitive reaction. In reality, a moral judgment only makes sense within the framework of the culture to which the subject belongs. This is because, as noted above, intuition is not a mere

[1] As I will argue later, this is an individualistic approach of moral judgment, which focuses on the subject who reacts passively to outside influences. Thus, Haidt opposes the rationalistic as well as individualistic approach, which makes judgment the result of individual reasoning.

automatic reaction carried out as a function of individual well-being, but as a function of the values that prevail in a culture that cultivates some intuitions and not others. Therefore, moral judgments should not be understood from an individualistic logic (the mere intuitive reaction of a subject to an external stimulus), but from a group perspective: the person values intuitively according to the intuitions culturally cultivated in their society. This explains Haidt's reference to Shweder's three cultures.

3.4. Moral reasoning

In the first part of the article, reasoning acquires an interpersonal characterization of an individualistic nature. Reasoning is a conscious, effortful and controlled process of transforming information to produce a moral judgment, not in oneself, but in another individual. Furthermore, reasoning acts in a biased way in the search for information. Concretely, reasoning occurs, when necessary, to justify a position previously adopted intuitively (Haidt 2001, 818).

With this characterization, reasoning is presented by Haidt as a *post hoc*, biased process that takes place in an interpersonal way, either to justify one's own position or to influence other people to generate in them the appropriate intuition. However, in this way, reasoning is presented in a framework of relationships between individuals, while again forgetting the social context surrounding them. Haidt does not explain, for example, why individuals need to justify their position, or what the subjects intend beyond changing the intuition of their dialogue partner. In other words, the exposition of the nature of reasoning relegates aspects that Haidt himself cites: for example, that relatedness motives presuppose a relationship between individuals and close subjects (and therefore, these motives presuppose the existence of a group to which the individual belongs); or the idea that biased reasoning seeks evidence that enables us to coincide with the members of our group, so that the collective maintains its stability (Haidt 2001, 826). Again, we see a contrast between an interpersonal reading which is individualistic in nature and an overlapping one which is group in nature.

3.5. Morality

In the first part of the article, Haidt characterizes moral evaluation as a process of intuition and perception, along the lines of Margolis: moral evaluation is, as noted above, an automatic evaluation of something as good or bad, but of whose origin the subject is not conscious (Haidt 2001, 822). However, in the second part, Haidt states that "morality is located in a group's efforts to solve cooperation and commitment problems" (Haidt 2001, 826). Again, we find a reading of morality in an individualistic logic (as an intuitive evaluation of the subject) in the first part, whereas in the second part morality makes sense in the practices of the group so that it can preserve its cohesion, and therefore morality has a group character.

Finally, the figure of the six processes that illustrates the social intuitionist model also helps reinforce the individualistic framework of interpretation of the social intuitionist model. Indeed, it depicts a subject A and a subject B interacting with each other, abstracted from any group context that might give meaning to their judgments.

Nonetheless, I do not claim that Haidt's exposition is contradictory. These different approaches to intuition, judgment, reasoning, morality and the social dimension, respond, as I have already argued, to the fact that the methodological approach to these concepts differs in each part of the work. Thus, whereas in the first part Haidt explains his model of social psychology in contrast to the rationalist theory, in the second part he tries to substantiate his proposal by supporting it with evidence from other disciplines. However, the key point here is that Haidt's argumentative strategy has far-reaching consequences for the interpretation of his theory:

- On the one hand, this sharp division between the two parts of the article makes Haidt's argumentation confusing and impedes an integrative understanding of the text. Furthermore, as I will argue below, the individualistic exposition in the first part of the article obscures the main ideas of the neuropsychological theory that Haidt outlines in this article and that he expands in later works.
- On the other hand, these features that define Haidt's argumentative strategy determine the reading of *The emotional dog and its rational tail* in the neurocentric paradigm. Unsurprisingly, the authors of this paradigm focus their criticisms on the theses set out in the first part of the article, while overlooking ideas that are outlined in the second part but that are key to giving true meaning to the social intuitionist model.[2] To demonstrate this point, in the following section I will present some of the criticisms made of the social intuitionist model in the neurocentric paradigm.

4. The reception of "The emotional dog and its rational tail" in the neurocentric paradigm

I call the "neurocentric paradigm" the neuroethical reflections derived from psychology and philosophy on the implications of the new theories of social psychology and evolutionary psychology regarding the nature of moral judgment and reasoning. This approach, developed from the 2000s to the middle of the 2010s, is centered on the psychological processes that take place in the minds of subjects and that give rise to moral judgment and reasoning. In this paradigm, studies have approached Haidt's article in relation to the proposals of other authors of neuroethics, such as Greene and Hauser.

[2] In particular, the neurocentric approach obviates the group dimension of the social intuitionist model. It is precisely this group dimension (developed by Haidt in his later works) that can in my opinion be rescued from the digital paradigm, as I will explain later.

I will divide my exposition into two parts. The first deals with the criticisms made of Haidt's article in the Anglo-Saxon world. The second part presents the criticisms made in the Spanish-speaking world, taking as a reference point those made by the Applied Ethics and Democracy Research Group (henceforth School of Valencia). This distinction is necessary, because the criticisms made in these two spheres, the Anglo-Saxon and the Hispanic, have relevant differences. I will begin by analyzing the reception of *The emotional dog and its rational tail* in the Anglo-Saxon sphere.

4.1. Criticisms of Haidt's article in the Anglo-Saxon world

The criticisms made of *The emotional dog and its rational tail* in the Anglo-Saxon world are framed within the analysis of theories of the dual process of moral judgment. Specifically, the reception of Haidt's article is framed within the reflections made on the role played by the two cognitive processes present in moral cognition: intuition and reasoning. For this reason, these criticisms of the text focus on Haidt's exposition in the first part of his article, in which he presents the elements that make up the social intuitionist model (intuition, judgment, and reasoning) and that distinguish it from the rationalist model. On this basis, the criticisms are characterized by two fundamental elements: they accept Haidt's thesis concerning the intuitive origin of most moral judgments, but at the same time they try to vindicate the role of reasoning in the formation of judgments. In what follows, I will outline the most representative criticisms.

The first criticism argues that Haidt underestimates the role of reasoning in the formation of judgments. Some authors (e.g., Fine 2006) argue that reasoning can break the connection between intuition and judgment, and thus prevent an intuitive judgment from being formed. For example, people do not apply stereotypes toward other groups when they recognize that doing so is inappropriate.

In a similar vein, other authors (e.g., Pizarro & Bloom 2003; Saltzstein & Kasachkoff 2004) argue that reasoning may not only reduce the power of intuition; it can also play a causal role in the formation of moral judgments in the subject, beyond the concrete situations recognized by Haidt (links 5 and 6 of the social intuitionist model). For example, as regards many current ethical problems, such as cloning, we do not have an intuitive answer prepared by evolution, and can only respond through reflection. Moreover, reasoning and intuition are connected, such that reasoning can modify what is intuitive to us at any given time. Thus, many of our current intuitive judgments may actually have a reflective origin, and through habit we have come to make them intuitively. For example, in ancient times it was intuitive to think that slavery was ethical, and it was hard to argue the immorality of slavery. However, today it is intuitive to think that slavery is unethical, because our culture has changed through the reflection of moral leaders who have brought about a change in values. Therefore, reasoning is essential to the formation of our moral judgments.

Third, Kennett and Fine (2009) defend that judgments made intuitively should not be recognized as having normative force if they differ from those that the subject would have made reflectively. That is, only those judgments that the subject forms through reasoning should count as proper moral judgments, because they are the only ones that the subject could consciously support. However, the latter idea is not adequate, because as the authors say in the text, following Saltzstein and Kasachkoff (2004) many of our intuitive judgments were originally formed reflexively, with habit leading us to make them intuitively. Therefore, there would also be intuitive judgments that the subject could recognize as their own. Here appears a constant in this Anglo-Saxon approach: Recognizing the intuitive origin of moral judgments encounters a dead end with any attempt to vindicate reasoning. This is the problem to which the School of Valencia seeks to find a solution.

Only one author, Steve Clarke (2008), has highlighted that the social dimension is completely absent from the discussion of Haidt's text. For this reason, a criticism that Clarke makes of Haidt is that his social intuitionist model may be valid for current forms of society in which intuition provokes automatic evaluations in individuals. However, perhaps it is possible to imagine other societies configured differently where intuition weighs less and people form their judgments through rational reflection. That is, the social intuitionist model cannot be applied in a timeless way to all cultures, but only to those existing now. Nevertheless, Clarke again assumes the intuitive origin of judgments and reduces the social dimension to interpersonal relationships.

Special mention should be made of another central author in the dual process model of moral judgment: Joshua Greene. This author is noteworthy because in his dialogue with Haidt's work we can find samples of the two paradigms that I present in this paper. In the decade after the publication of *The emotional dog and its rational tail*, Greene criticized the excessive weight given by Haidt to intuition in the formation of moral judgment. In fact, he came to consider Haidt a purely emotivist author as opposed to merely another member of the dual process model, precisely because he does not recognize reasoning as the cause of many of our moral judgments (Greene 2008; Paxton & Greene 2010). In his work, Greene defends the dual process model of moral judgment in which reasoning gives rise to utilitarian judgments, while arguing that the intuitive process produces deontological judgments. However, in his book *Moral tribes* Greene makes a reconsideration of Haidt's work using the holistic or group approach. This group approach is the key element of what I call the digital paradigm. I will refer to this later.

Let us now move on to analyze the reception of Haidt's text in the Hispanic literature.

4.2. Criticisms of Haidt's article in the Hispanic world

In the Hispanic world, the main criticisms of Haidt's article have been made by the School of Valencia, directed by Adela Cortina and of which I am a member. This group's analysis

of Haidt's work has been carried out within the field of neuroethics in a series of research projects developed in the 2000s and 2010s. In this sense, the School of Valencia's work is also neurocentric. The fundamental difference between the approach of the School of Valencia and the Anglo-Saxon approach lies in the fact that the former does not limit itself to defending a greater weight of reasoning in the formation of moral judgments. The central argument of the School of Valencia is that the analysis of Anglo-Saxon neuroethics on the origin of moral judgment is reductionist, because, as mentioned above, it is limited to gauging the real weight of the different cognitive processes (intuition and reasoning) in the formation of judgments. In contrast to the Anglo-Saxon approach, the School of Valencia argues that moral judgment is a holistic process that integrates various elements, psychological processes being only one of them. Cortina and I are the researchers at the School of Valencia who have approached Haidt's work in the most detail. I will begin by addressing the main criticisms made by Cortina, and then outline my own.

Cortina has developed her study of neuroethics from her notion of cordial reason. She has been developing this concept since the beginning of the 2000s, taking clear shape in her work Ética *de la razón cordial* (2007). It was within the framework of this ethical proposal, which seeks to integrate emotion within rationality, that Cortina fully entered the discussion on neuroethics and neuropolitics. As a result of her work in these disciplines, she published a series of works, the most important being *Neuroética y neuropolitítica* (2011). Cortina has made several criticisms of the neuroscientific approach to morality, but here I will solely rough out the three main ones that affect the theory of the social intuitionist model.

The first criticism refers to the very nature of the dual process model. According to Cortina, affirming that intuition and reasoning are distinct psychological processes with distinct brain bases and a different evolutionary origin, as claimed by Greene and Haidt, has serious consequences for our conception of ourselves as moral beings. According to Haidt's social intuitionist model, subjects are aware of the judgments they make, but not of the reasons why they make them. Moreover, the reasons they express to justify their judgments derive from a confabulation process that reproduces reasons learned in society. Cortina argues that this situation inescapably turns the subjects into victims of a form of "moral schizophrenia" that annuls their moral agency (Cortina 2011). However, this is a conception far removed from the awareness we have of ourselves as moral subjects and agents with moral autonomy. This supposed moral schizophrenia can be perfectly explained, according to Cortina (2011), by appealing to the heteronomous character of many of our judgments, without this nullifying our moral agency. This brings us to the second criticism.

The moral schizophrenia to which Haidt's social intuitionist model leads is the result of the myopic approach that neuroethics adopts on moral judgment. For Cortina (2011), the authors of this discipline tend to confuse the *bases* of moral judgment with its *foundations*. When making moral judgments, there is no doubt that a series of psychological processes are produced and specific regions are activated in our brains. These processes are a necessary *basis* for forming these judgments, because without them we would not be able to make moral

judgments. However, psychological processes do not provide the *foundation* for our moral judgments. The foundation falls into the properly moral sphere, related to a demand for universality that goes beyond the adaptive codes that evolution has embedded in our brains. That is, we do not act morally because an intuition is activated or because the judgment is the result of a reasoning process, but because we recognize others as equal beings in dignity and we have to give reasons for our actions.

Finally, Cortina (2011) criticizes the methodologies of Greene and Haidt. Cortina points out that these authors' theses are based on experiments in which the variables that the experimenter wants to take into account are artificially limited: Neither Haidt's extravagant situations nor Greene's dilemmatic situations respond to the moral situations that people face in real life. People face problems, not dilemmas. Dilemmas present a tragic situation in which alternatives are given beforehand and where no option is morally better than another. By contrast, problems are complex situations with an open-ended solution that subjects have to find by taking various factors into account.

For my part, I began to study neuroethics as a postdoctoral research fellow at the *Uehiro Centre for Practical Ethics* at the University of Oxford. There I focused my research on the theories of Greene and especially Haidt on the sources of moral judgment. I also participated in a reading group directed by Steve Clarke on Haidt's book *The righteous mind*. Next, as a lecturer in Philosophy at the University of Valencia, I extended my analysis to the field of neuropolitics. My criticisms of Haidt focus on the model of public deliberation articulated in the social intuitionist model. I have two main criticisms, which I will share next.

In line with Cortina, I argue that we cannot reduce moral judgment to a concatenation of psychological processes, some conscious and others unconscious (Pérez Zafrilla 2013). This reductionist view forgets that moral judgment is a complex process of a constitutively reflexive nature that, as Aristotle long ago demonstrated, combines several elements. First, there is an evaluation of circumstances and a search for solutions. Second, there is an attribution of intentionality. Indeed, only if we are able to attribute intentionality to the subject can we say whether their action is moral or not. However, we must also bear in mind that many of our judgments are heteronomous, and that this heteronomy explains the phenomenon of moral dumbfounding, without the need to postulate the intuitive origin of judgments, as Haidt and his Anglo-Saxon critics believe. Finally, moral judgment makes a pretension to universality (what is contrary to what is affirmed as just is not considered acceptable) and a pretension to convince others with reasons (it is hoped that the judgment will be accepted by all as just).

The second criticism pertains to the model of public deliberation in the social intuitionist model. Here again it is a mistake to reduce public deliberation to a struggle to create new intuitions in the interlocutor, according to the circle of links 3, 1 and 2 of the social intuitionist model. My critique is primarily aimed at dismantling the illusions of moral deliberation cited by Haidt. The first illusion is the belief that judgments proceed from a reasoning that objectively evaluates reality, when in fact judgments, for Haidt, are intuitive. The second

illusion is the belief that the exchange of reasons manages to convince the interlocutor. According to Haidt, what the argumentative exchange actually achieves is the polarization of the dialogue, because of the generation of reactive feelings of suspicion and frustration at not being able to convince the interlocutor.

In my opinion, both illusions respond to a reductionist conception of moral deliberation which is unable to distinguish the different forms of communication highlighted by Habermas (1984): strategic action (guided by the imposition of interests) and communicative action (in which the subjects are guided by a criterion of justice and seek rational agreement). In the first form of communication, judgments express selfish preferences, while in the second, judgments express claims of justice. Moral deliberation corresponds to this second form of communication, while Haidt's analysis is closer to strategic action. This is so because moral deliberation is based on cognitive presuppositions set forth by deliberative democracy, such as those mentioned above: Judgments express a pretension to universality and to convince others with reasons. However, deliberation is also based on basic moral principles: the recognition of the other as a valid interlocutor (Cortina 2007), the symmetry of the parts, being guided by a criterion of justice (and not by selfish interests) or the use of reasons that the other can accept, and the exclusion of other forms of communication, such as fallacies or demagogy. For this reason, the expectation to convince others with reasons, far from being an illusion, is a constitutive element of public deliberation (Pérez Zafrilla 2017a).

In the same way, the reactive feelings of frustration and suspicion toward the interlocutor that arise in the deliberative process may have an alternative explanation to that highlighted by Haidt. My thesis is that these feelings arise for two reasons: either because subject A recognizes that their interlocutor B is not behaving in an ethical manner in the dialogue (when B appeals to demagogy and does not recognize A as a valid interlocutor); or because agreement is impossible as both interlocutors appeal to incommensurable values (Pérez Zafrilla 2017b). In any case, and this is the key point, such feelings of frustration, indignation and suspicion toward the interlocutor, far from revealing the intuitive origin of the judgments, have a purely moral basis.[3] Indeed, these moral feelings alert the subject that the interlocutor does not have an ethical attitude (by appealing to demagogy) or holds values that the subject considers immoral. In this sense, too, the polarization produced in dialogue as a result of the appearance of these reactive feelings, far from being a pathology of deliberation, is proof that dialogue is governed by ethical principles that must be respected (Pérez Zafrilla 2017b).

Therefore, the School of Valencia separates itself from the Anglo-Saxon approach regarding two points. First, the School of Valencia denounces the psychologistic reductionism of the Anglo-Saxon approach, which focuses on the weight of cognitive processes (intuition and reasoning) in the formation of moral judgment. Second, the School of Valencia stresses that moral judgment and moral deliberation constitute holistic processes in which psychological

[3] This inability of Haidt to recognize the moral character of certain emotions is another failure of the social intuitionist model, pointed out by Cortina with her ethics of cordial reason.

processes are just one factor among others cited above (i.e., attribution of intentionality), but are not the proper moral element. However, like the Anglo-Saxon model, the School of Valencia focuses its criticisms on the individualistic reading present in the first part of Haidt's article. Cortina and I criticize the reduction of morality or deliberation to psychological processes, but we approach intuitions, reasoning, judgments and morality (with the pretensions of universality of moral judgments) as processes that occur in the minds of interacting subjects. Certainly, the School of Valencia maintains that judgments are formed in a context. Furthermore, Cortina asserts that moral emotions are cultivated, through education, in relation to other subjects (Cortina 2007). However, the point here is that these criticisms are made by obviating the group framework of interpretation (or holistic reading) that can be made of Haidt's article. Therefore, the neurocentric approach of the School of Valencia remains, like the Anglo-Saxon one, in line with Haidt's individualistic reading.

For all these reasons, in order to make a reading of *The emotional dog and its rational tail* that integrates the elements of the social intuitionist model in a social context, we should abandon the neurocentric approach, centered on analyzing the psychological processes that occur in subjects' minds, and adopt a new, broader approach that conceives individuals as members of a social reality. This is precisely what I aim to do now within the School of Valencia. I underpin a new approach to Haidt's article on the new paradigm of study that has emerged in recent years thanks to the development of social networks.

5. The reception of "The emotional dog and its rational tail" in the digital paradigm

I use the phrase "digital paradigm" to refer to the current reflections made in academia on the impacts of new technologies, social networks and artificial intelligence on our understanding of human beings. The main change brought about by social networks is that the subject can no longer be conceptualized as self-centered. On the contrary, the individual lives projected toward their environment in search of recognition and attention, as Byung-Chul Han (2017) argues. My thesis is that this new approach toward the external image enables a new reading of *The emotional dog and its rational tai*l to be articulated. Even more, this new reading is the key to recognizing in Haidt's article of 2001 how psychological processes, exposed in the first part of the text, make real sense in the social sphere. However, above all, this new approach represents a Copernican turn in Haidt's study. In light of what I have discussed in the previous sections, the neurocentric approach is interior oriented: It analyzes how phenomena occurring externally trigger cognitive processes of an unconscious and reactive nature in the individual's mind. By contrast, the digital paradigm is exterior oriented: It approaches psychological processes as adaptive mechanisms that enable the subject to project an image of themselves to the other members of the group to which they belong. That is, whereas the neurocentric approach goes from the outside in, the digital paradigm goes from the inside out, projecting the individual in a group context. This new approach will make it possible to better integrate Haidt's article into his later work as a whole.

There are four pillars that support this new approach to *The emotional dog and its rational tail*. The first pillar comprises the theses maintained by Haidt in later works, mainly in *The righteous mind*. The second pillar is Haidt's recognition of theories of group selection, from which he develops his conception of morality. The third pillar corresponds to new theories of moral psychology that have recently emerged, which emphasize the social character of judgments and moral reasoning. The final pillar is the technological revolution that has arisen with social networks, because the use of these technologies exponentially increases the effect of the biases inherited from evolution that reinforce our tribal nature. Due to space limitations, I will confine myself to briefly outlining each of these points.

First, a holistic reading of *The emotional dog and its rational tail* is possible by attending to later developments in Haidt's work. Of particular interest here is not so much the theory of moral foundations, but the synthesis of moral psychology that Haidt establishes in various works. Based on studies carried out in evolutionary psychology in the preceding decades, Haidt (2007; Haidt & Kesebir 2010) argues that moral psychology can be structured around three principles: intuitive primacy (but not dictatorship); moral thinking is for social doing; and morality binds and builds.[4] It is easy to see that the individualistic reading of the social intuitionist model (centered on the psychological processes occurring in the mind of the subject) only makes sense within the first of these principles. However, Haidt's individualistic reading makes the social intuitionist model completely alien to the other two principles of moral psychology.

Instead, a group approach to *The emotional dog and its rational tail*, typical of the digital paradigm, enables us to integrate this text into the other two principles as well. Indeed, intuitions, judgments and reasoning only make sense if we understand them as instruments for subjects to function effectively in a social environment. For example, when Haidt states in *The emotional dog and its rational tail* that "moral reasoning is produced and sent forth verbally to justify one's already made judgment to others" (Haidt 2011, 818–819), he is not merely saying that reasoning is there to provide justifications for our actions (because reasoning has a biased and *a posteriori* functioning), as interpreted in the neurocentric paradigm, focused on the analysis of psychological processes. In reality, this phrase only makes sense in conjunction with the other two principles of social psychology. Thus, the principle of "moral thinking is for social doing" teaches us that subjects justify their actions in order to maintain their reputation. Furthermore, subjects seek to generate intuitions in other subjects in order to project a good image to them. In this way, the psychological processes that define the social intuitionist model make sense in the fact that the subject lives with an eye to the outside world. In the same way, "morality binds and blinds" because intuitions are not mere affective responses provoked in our minds by the environment, but responses cultivated in some

[4] In *The righteous mind* these principles are articulated as: intuitions come first, strategic reasoning second; there's more to morality than harm and fairness; and morality binds and blinds.

groups according to the virtues they foster in order to maintain cooperation. Therefore, it is in the framework of group practices that the social intuitionist model makes sense, not in the minds of the subjects.

Second, this Copernican turn in the reading of the social intuitionist model, which makes it possible to integrate the 2001 text into Haidt's work as a whole, is supported by the same works of evolutionary psychology that Haidt takes as referents for his proposal. Haidt adopts a conceptualization of morality understood as "interlocking sets of values, virtues, norms, practices, identities, institutions, technologies, and evolved psychological mechanisms that work together to suppress or regulate self-interest and make cooperative societies possible" (Haidt 2012, 270). This Durkheimian model is based on the theories of evolutionary psychologists and primatologists related to the theory of group-level selection. Prominent among these are Darwin's (2017) theory of the evolution of social sentiments, Alexander's (1986) theory of indirect reciprocity, and Dunbar's (1996) theory of language evolution. Again, all these theories establish a Glauconian social model in which subjects live concerned about their reputation before others, this being the best way to maintain their survival in the group. Nevertheless, this individual search for reputation produces more cohesive groups that are able to prevail over less cohesive groups in a group-level selection.

Third, this digital paradigm is supported by the innovative theories of evolutionary psychology that have recently appeared and that Haidt also takes as referents for his theory. Among them, the argumentative theory of reasoning stands out. The core thesis of this theory is that the main function of reasoning is not cognitive, but argumentative (Mercier & Sperber 2011). That is, reasoning has not evolved because it enables us to know the world, but because of its advantages in facilitating the survival of subjects in the group, seeking and exposing the appropriate arguments that enhance subjects' reputation in the eyes of others. Subjects, precisely because they are social beings, use argumentation to project to others a good image of themselves that will enable them to improve their status in the group and guarantee their survival. This explains the existence of biases such as confirmation bias or motivated reasoning. In everyday life, reasoning does not seek to have an objective view of reality. It proceeds in a biased way, seeking and giving more weight to evidence that favors our position or that of those close to us (our party, for example), because this is what enabled our ancestors to survive in group environments. Therefore, these biases, far from being a programming error of our rational mind that we should correct, constitute a function of our argumentative mind in a social context. This is because the social dimension of reasoning (the social consequences of what subjects say in public) undermines the capacity of our faculty of reason to escape these conditioning factors. In fact, the motivation guiding subjects' behavior in the group is precisely their desire to maintain their reputation (Mercier 2011; Pérez Zafrilla 2016). This argumentative theory of reasoning also gives meaning to the metaphor used by Haidt that reasoning behaving is more like a lawyer defending their client than a judge trying to seek the truth.

Greene should be mentioned again here. In *Moral tribes* Greene develops ideas already implicit in his previous papers concerning the tribal nature of our brain configuration. The brain evolved to favor collaboration with those close to us and to promote competition with other groups. In this sense and in line with Haidt, Greene understands morality as a biological adaptation to promote cooperation and restrict selfishness within the group. Thus, subjects are tribal beings who seek to maintain their reputation by cooperating within the group or by confabulating stories that can enhance their position within it. Thus, like Haidt, Greene in *Moral tribes* reorients his analysis of morality from a neurocentric approach to a group approach.

The final pillar articulating this digital paradigm is the configuration of the new digital environment with the popularization of social networks. When the Internet first emerged, there was hope of creating a digital public sphere open to dialogue among different people (Carr 2011). However, as Haidt and Rose-Stockwell (2019) note, over time and especially with the popularization of social networks, a completely different digital environment has developed in which emotionality and polarization prevail. An explanation of this contrast between initial expectations and the subsequent development of this digital environment can be found in the reflections on the digital world made by some authors. A clue exists in the important distinction made by Han (2017) between analogue and digital media: Analogue media (television, radio) group individuals by diluting them into a mass, whereas digital media (the Internet and social networks) isolate individuals in their rooms. In this way, the digital medium turns subjects into individuals in need of attention. This isolation stimulates subjects to go out into cyberspace to look for other peers with whom to satisfy their affective needs for recognition and belonging. For this reason, in social networks, expressive communication prevails over rational dialogue. People use social networks not to engage in dialogue with others, but to express opinions, feelings and moods, and they do so with emotionally charged messages that attract the attention of like-minded people (Arias Maldonado 2016).

This primacy of emotionality and the cultivation of our external image in the digital sphere is also expressed, paradoxically, through the use of anonymous profiles. Many people choose such profiles to adopt a group identity, activating what Haidt calls the "hive switch": the ability to transcend self-interest and to think and feel as members of a group (Haidt 2012). In the framework of group identity, the subject tries to maintain a positive image of the group in the face of adversaries, while additionally attempting to preserve their personal reputation within the group (Brady et al. 2020). To achieve these goals, people convey emotions of indignation toward adversaries, because as Brady et al. (2020) argue, expressing emotions such as anger or indignation can enable them to go viral on the network. Indeed, communicating such emotions enables subjects to reinforce their reputation within the group, thanks to features of social networks such as likes, followers and retweets, which reward them for publishing content that strengthens the group's position as a whole. Such validation by the group, in turn, encourages subjects to publish more radical content in order to obtain more notoriety and further improve their position in the group.

For all these reasons, the digital world is configured as an environment in which individuals live outwardly oriented, seeking approval and recognition from others and exhibiting moral indignation that enhances their reputation in the group. Moreover, just like in real life, biases that remind us of our tribal nature emerge in this digital world, too. In fact, in the digital environment biases appear more frequently, as social networks act as a supermoral stimulus (Crocket 2017). That is, social networks break the relationship between distant and near people. On the Internet, anyone is close, even where they are physically thousands of kilometers away. Furthermore, networks provide us with more occasions to be outraged, because the Internet exposes us to a large number of immoral acts.

In this sense, recent reflections in philosophy and sociology on the influence of the digital environment on our lives, in relation to issues such as moral grandstanding (Tosi & Warmke 2016), virtue signaling (Miller 2019), firestorms (Han 2017), the challenges of privacy and artificial intelligence (Véliz 2020), and artificial polarization on the Internet (Pérez Zafrilla 2021), make up a digital paradigm that is quite different from the neurocentric paradigm of the 2000s. This digital paradigm regards subjects as members of a group, and focuses on the image that individuals project to the outside world. For this reason, this paradigm of reflection is appropriate terrain for returning to Haidt's (2001) article and reading it in such a way that integrates it within the rest of this author's work.

6. Conclusion

Twenty years after the publication of *The emotional dog and its rational tail*, we can return to Haidt's seminal article with a certain perspective in order to analyze the theoretical reception of the text. As I have argued, the study of the article has been conditioned by the prevailing theoretical paradigm of the time. During the first period, the text was read from a neurocentric paradigm, addressing the development of neuroethics and the implications of neuroscientific advances and the dual process model on our conception of morality. However, this approach turned out to be myopic. For instance, it was unable to analyze the various psychological processes in the context surrounding the individual in order to recognize their true meaning. Furthermore, it introduced a break in the reception of Haidt's work: The social intuitionist model was disconnected from this author's later writing, centered on the theory of moral foundations and his Durkheimian model of morality. The model had an individualistic nature, in contrast to other proposals of a clearly group-oriented nature.

However, the current digital paradigm enables us to return to the text with a new perspective. The focus today is no longer on the psychological processes taking place in the individual's mind, but on the subject's behavior with the aim of maintaining their reputation in the (digital) environment, as they are now seen as a member of a group. This reading is particularly appropiate as it allows us to integrate the social intuitionist model into Haidt's production as a whole. Therefore, we can say that the emotional dog we all knew two decades ago was a Glauconian canine, which today is perfectly integrated into the digital environment.

Acknowledgment

This publication has been supported by the Scientific Research and Development Project PID2019-109078RB-C22 funded by MCIN/ AEI /10.13039/501100011033.

References

Alexander, R. (1987). *The biology of moral systems.* New York: Routledge.

Arias Maldonado, M. (2016). *La democracia sentimental. Política y emociones en el siglo XXI.* Barcelona: Página indómita.

Brady, W., Crockett, M. J., Van Bavel, J. (2020). The MAD Model of Moral Contagion: The role of Motivation, Attention and Design in the spread of moralized content online. *Perspectives on Psychological Science*, *15*, 978-1010.

Carr, N. (2010). *The shallows: what the Internet is doing to our brains*. New York: W.W. Norton.

Clarke, S. (2008). SIM and the city: rationalism in psychology and philosophy and Haidt's account of moral judgment. *Philosophical Psychology*, *21*(6), 799-820.

Cortina, A. (2007). Ética de la razón cordial. Oviedo: Nobel.

Cortina, A. (2011). *Neuroética y neuropolítica*. Madrid: Tecnos.

Crockett, M. (2017). Moral outrage in the digital age. *Nature Human Behavior*, *1*, 769-771.

Darwin, C. (2017). *The descent of man*. London: Penguin Classics.

Dunbar, R. (1996). *Grooming, gossip and the evolution of language*. Cambridge: Harvard University Press.

Fine, C. (2006). Is the emotional dog wagging its rational tail, or chasing it? *Philosophical Explorations*, *9*(1), 83-98.

Greene, J. (2008). The secret joke of Kant's soul. In Walter Sinnott-Armstrong (ed.), *Moral Psychology Vol.3*, pp. 38-79. Cambridge: MIT Press.

Greene, J. (2013). *Moral tribes. Emotion, reason, and the gap between us and them*. London: Atlantic Books.

Habermas, J. (1984). *The theory of communicative action*. Boston: Beacom Press.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.

Haidt, J. (2007). The New Synthesis in Moral Psychology. *Science*, *316*, 998-1002.

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

Haidt, J., Kesebir, S. (2010). Morality. En S. Fiske, D. Gilbert y G. Lindzey (eds.). *Handbook of social Psychology*, pp. 797-832. Hobeken (NJ): Wiley.

Haidt, J., Rose-Stockwell, T. (2019). The dark psychology of social networks. Why it feels like everything is going haywire. *The Atlantic*, December. Last access 29th march 2021. Available in: https://www.theatlantic.com/magazine/archive/2019/12/social-media-democracy/600763/

Han, B.-C. (2017). *In the swarm: digital prospects*. Cambridge: The MIT Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Guiroux.

Kennett, J., Fine, C. (2009). Will the real moral judgment please stand up? *Ethical theory and moral practice*, *12*(1), 77-96.

Margolis, H. (1987). *Patterns, thinking and cognition.* Chicago: University of Chicago Press.

Mercier, H. (2011). What good is moral reasoning? *Mind & Society*, *10*, 131-148.

Mercier, H., Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory, *Behavioral and Brain Sciences*, 34, 57-111.

Miller, G. (2019). *Virtue Signaling: Essays on Darwinian Politics & Free Speech*. s.l.: Cambrian Moon.

Nisbett, R. E., Wilson, T. (1977). Telling more than we can know: verbal reports on mental process, *Psychological Review*, 84, 231-259.

Paxton, J., Greene, J. (2010). Moral reasoning: Hints and Allegations. *Topics in Cognitive Scienc*e, *2*(3), 511-527.

Pérez Zafrilla, P. J. (2013). Implicaciones normativas de la psicología moral: Jonathan Haidt y el desconcierto moral. *Daimon*, *59*, 9-25.

Pérez Zafrilla, P. J. (2016). Is Deliberative Democracy an adaptive political theory? A critical analysis of Hugo Mercier's Argumentative Theory of Reasoning. *Análise Social*, *51*(3), 542-562.

Pérez Zafrilla, P. J. (2017a). Illusions and reality of public deliberation, *Cogency*, *9*(1), 53-72.

Pérez Zafrilla, P. J. (2017b). Por qué fracasa la deliberación y cómo podemos remediarlo. Una alternativa ética al enfoque neurocientífico. *Daimon. Revista Internacional de Filosofía*, 70, 131-146.

Pérez Zafrilla, P. J. (2021). Polarización artificial: cómo los discursos expresivos inflaman la percepción de polarización política en internet. *Recerca. Revista de pensament i anàlisi*, 22, in press.

Pizarro, D. A., Bloom, P. (2003). The intelligence of the Moral intuitions: comment on Haidt (2001). *Psychological Review*, *110*(1), 193-196.

Roskies, A. (2002). Neuroethics for the new millenium. *Neuron*, *35*, 21-23.

Saltzstein, H. D. y Kasachkoff, T. (2004). Haidt's moral intuitionist theory: A psychological and philosophical critique. *Review of General Psychology*, *8*(4), 273-282.

Tosi, J., Warmke, B. (2016). Moral grandstanding. *Philosophy & Public Affairs*, *44*(3), 197-217.

Véliz, C. (2020). *Privacy is power.* London: Transword.

Zajonc, R.B. (1980). Feeling and thinking: preferences need no inferences. *American Psychologist*, *35*, 151-175.