Intuitionism, Practical Reasoning and Defeasibility

DANIEL **GONZÁLEZ LAGIER**
University of Alicante

ABSTRACT
This text considers the contributions that cognitive sciences can make to the study of legal reasoning, distinguishing between descriptive, conceptual and normative impacts. In particular, it is concerned with exploring Jonathan Haidt's social intuitionism thesis, which says that, when we reason about moral and practical issues, we make a decision intuitively, which we then rationalise *a posteriori* (although at this stage reason cannot change the decision made). The text considers how this thesis would apply to the problem of the defeasibility of rules, which it takes as one of the characteristic features of legal reasoning. Finally, some objections are presented to Haidt's thesis and to the normative claims of some cognitive scientists.

KEYWORDS
cognitive sciences, social intuitionism, defeasibility, what is and what should be

CITATION
González Lagier D. 2023. *Intuitionism, Practical Reasoning and Defeasibility*, in Brigaglia M., Roversi C. (eds.), *Legal Reasoning and Cognitive Science: Topics and Perspectives*, «Diritto & Questioni pubbliche», Special Publication, August 2023.

DIRITTO & QUESTIONI PUBBLICHE / RECOGNISE
Legal Reasoning and Cognitive Science: Topics and Perspectives
Special Publication / August, 2023, pp. 389-404

ISSN 1825-0173

Intuitionism, Practical Reasoning and Defeasibility

DANIEL GONZÁLEZ LAGIER

1. *The relevance of cognitive sciences for practical reasoning: some examples* – 2. *The "intuitionist" model of normative reasoning* – 3. *The defeasibility of rules and the intuitionist model* – 4. *Some doubts about the plausibility of the intuitionist model: descriptive claims* – 5. *The problem of justification and normativity* – 5.1. *Normative claims based on the cognitive sciences* – 5.2. *Is it possible to infer normative conclusions from purely descriptive statements? Arguments against Hume's Law*

1. *The relevance of cognitive sciences for practical reasoning: some examples*

The term "cognitive sciences" usually refers to a set of studies (stemming from psychology, neuroscience, artificial intelligence, linguistics, cognitive anthropology, and so on) that, from an experimental and scientific point of view, deal with the mind and the capacities related to the acquisition of knowledge and decision making. The cognitive sciences have recently undergone enormous development, largely supported by neuroscientific techniques for observing brain function. These disciplines can shed much light on important aspects of practical reasoning in general, and legal reasoning in particular, as their contributions show general characteristics of human thought processes that are clearly reflected in the latter. The following are examples of the importance they may have for, on one hand, establishing the possibility of human rationality (and the type of rationality) and, on the other, concerning discussions on the theory of law and legal argument:

a) Kahneman and Tversky have shown that human reasoning proceeds by means of heuristics and shortcuts that lead to numerous biases and errors (KAHNEMAN 2011, KAHNEMAN 2003). These biases lead us away from the answers that would be correct in accordance with a standard conception of rationality based on deductive, inductive and probabilistic rules of inference. This type of research therefore appears to assume a pessimistic conception of human rationality (which they call limited rationality). We simply cannot know whether a belief of ours is true or false, or whether it is based on correct reasons, as biases are usually unconscious. Legal reasoning would not escape these biases, which would affect all areas, from the legislative to the judicial. For example, studies have been made of how the way the facts in the prosecution's indictment are classified creates an *anchoring bias* in the judge (FARIÑA et al. 2002), so it would be necessary to consider whether and how these can be overcome (institutional modifications, argumentation strategies, etc.). The study of biases in legal reasoning could be seen as a complement, from a psychological perspective, to the traditional analysis of fallacies carried out by the theory of argumentation.
b) Overcoming the pessimism of bias-based studies, evolutionary psychology has argued that the human mind consists of a set of information processing and decision-making modules or systems that have been designed by natural selection. This ensures that human beings are rational, at least in the sense that their decisions and actions are (considered as a whole) instrumentally adequate to meet goals relevant to adaptation and survival, i.e. biological goals (GARCÍA CAMPOS 2011). A vision similar to this is behind the authors who have been constructing new disciplines, such as "neuroethics" and "neurolaw", which—with the support of neuroscientific research—argue that moral reasoning (and legal reasoning, to the extent that it is permeated by moral considerations) is determined by intuitions and emotions formed by a process of natural selection. In other words, they have an important

adaptive value for human beings (for example, we could say that we believe solidarity with the members of our community is an important moral value because supportive and cooperative behaviour favours the survival of the species) (González Lagier 2017). Marc Hauser, for example, has even postulated the existence of a moral module or organ that is decisive in our deliberation when we are faced with moral dilemmas (Hauser 2006).

c) However, the notion of rationality of evolutionary psychology and neuroscience is not the same as the "tradition of biases and heuristics". It could be said that evolutionists assume an instrumental or consequentialist notion of rationality based on efficiency for survival, while Kahneman, Tversky and many others assume a deontological conception based on rules or standards (of deductive and inductive logic). Is there a way to accommodate both types of rationality in human behaviour? It has been suggested that "dual systems" or "dual processing" theory attempts precisely this type of reconciliation (García Campos 2012). To this end, the theory postulates that human beings have two reasoning or decision-making systems: one of them is fast, intuitive, non-verbal and unconscious (we are aware only of the result of reasoning), evolutionarily older and shared with other animals. Following the convenient terminology used by Kahneman, we can call this System 1. The other is slow, intentional, methodical, conscious, verbal (or verbalisable), developed later and characteristic of human beings (System 2) (Kahneman 2011). The first system would correspond to a strategic rationality – the result of natural selection – while the second would correspond to rule-based rationality, more connected to the social and cultural dimension of human beings (García Campos 2009, 78). What is rational in one sense need not coincide with what is rational in the other, so there may be a conflict between the two systems. Obviously, this discussion is relevant to understanding how legal reasoning works and how the two types of decision processing come together: For example, Manuel Atienza has suggested that the articulation between System 1 (the intuitive) and System 2 (the slow and rational) could account for the interplay between rules (which would guide quick and intuitive decisions) and principles (whose application would require more complex reasoning) (Atienza 2017, 429).

d) Even when a dual conception of reasoning that leaves room for reason (in the sense of deductive, inductive procedures, etc.) is maintained, it will not necessarily play an important role in the correctness and justification of decisions. For example, Jonathan Haidt, in the view of practical reasoning that he calls "social intuitionism", accepts (as we shall see below) the thesis of the two systems of decision-making, but considers that System 2 (the rational one) does not really fulfil a control function for System 1, at least when it comes to making moral decisions. In most cases, in his view, System 2 performs only a function of a *posteriori* rationalisation of intuitively made decisions, without attempting to change them:

> «Once people find evidence to support them, even a single piece of bad evidence, they often stop the search, since they have a "make-sense epistemology" [...] in which the goal of thinking is not to reach the most accurate conclusion but to find the first conclusion that hangs together well and that fits with one's important prior beliefs» (Haidt 2001, 821).

If Haidt is right, the distinction between context of discovery and context of justification, with which an attempt has been made to refute the objections of sceptics about the justification or correctness of judicial decisions, is no longer important. It can no longer be said that what matters in fact is not how a conclusion is reached, but whether it is justified in the light of rational considerations because these would not be oriented towards correctness, but rather motivated by consistency with the results of System 1, whatever they may be.

In general, three levels or areas of possible relevance of cognitive sciences can be distinguished: descriptive, conceptual and normative. At the first level, cognitive sciences would provide a description of how, in fact, reasoning and decision-making processes take place in the

mind (e.g., which brain modules are involved in moral problem solving). We could say that, insofar as these are empirical disciplines, this is their most natural function. At a conceptual level, cognitive sciences can provide knowledge to be taken into account in the construction of concepts hitherto considered to be philosophical, such as the mind, rationality, reasoning, decision, volition, emotion, and so on. At a normative level, they may help to construct criteria for evaluating different reasoning strategies and suggest which ones are more appropriate in relation to certain purposes, i.e., in some sense they may help to choose which is the best strategy for reaching a rational justified decision (e.g., some psychologists suggest that probabilistic reasoning is improved if problems are presented in frequency format rather than in percentage format, from which it would follow that they *should be* presented in frequency format) (García Campos 2014, 26). Clearly, it is at the conceptual and normative level that there is the greatest interconnection between cognitive science and philosophy. And it is the possibility that the cognitive sciences have something to say (and the extent to which this is the case) at the normative level that generates the most controversy. A common thesis is that a distinction must be made between the explanation and the justification of a decision or a belief. While cognitive sciences can be very useful in providing explanations of our way of reasoning and deciding, by themselves they can contribute little to its justification, because this would imply moving from facts to rules, which would be forbidden by several arguments (such as George Moore's naturalistic fallacy argument and "Hume's guillotine").

In this paper I will try to discuss the plausibility, for legal reasoning, of the intuitionist model of moral judgement put forward by the psychologist Jonathan Haidt. Of course, it is not the only model proposed from the cognitive sciences, nor the only one that resorts to intuitions to explain decision-making, but it is one of the most influential and well-known in moral and legal philosophy (although Haidt is thinking primarily of moral reasoning, it is not difficult to project his ideas on to legal reasoning). In particular, I will take as a "test bed" the problem of the defeasibility of rules. To do so, I will proceed as follows: firstly, I will present the intuitionist model; secondly, I will briefly explain what the defeasible nature of legal reasoning consists of and how it could be explained by the intuitionist model. I will then make some critical remarks about this model. Finally, I will discuss whether a model of reasoning based on cognitive science can have a normative scope, in the strong sense. Whether, as well as having explanatory relevance, it can also provide guidelines for the justification or correctness of normative reasoning.

2. *The "intuitionist" model of normative reasoning*

In his paper *The emotional dog and its rational tail: a social intuitionist approach to moral judgment*, Jonathan Haidt proposes a model of explanation of practical reasoning that he calls "social intuitionism": intuitionism because of the role it gives to intuitions and emotions, and social intuitionism because of the role it gives to the influence of social and cultural aspects.

The main thesis of social intuitionism is, according to Haidt himself, «that moral judgment is caused by quick moral intuitions, and is followed (when needed) by slow, ex-post facto moral reasoning» (Haidt 2001, 817). The author illustrates his thesis with an experiment and justifies it with several indications of the causal ineffectiveness of reason for the formation of moral judgement. The experiment is as follows:

> «"Julie and Mark are brother and sister. They are traveling together in France on summer vacation from college. One night they are staying alone in a cabin near the beach. They decide that it would be interesting and fun if they tried making love. At the very least it would be a new experience for each

> of them. Julie was already taking birth control pills, but Mark uses a condom too, just to be safe. They both enjoy making love, but they decide not to do it again. They keep that night as a special secret, which makes them feel even closer to each other. What do you think about that? Was it OK for them to make love?"».
>
> «Most people who hear the above story immediately say that it was wrong for the siblings to make love, and they then begin searching for reasons [...]. They point out the dangers of inbreeding, only to remember that Julie and Mark used two forms of birth control. They argue that Julie and Mark will be hurt, perhaps emotionally, even though the story makes it clear that no harm befell them. Eventually, many people say something like "I don't know, I can't explain it, I just know it's wrong"» (HAIDT 2001, 814).

Haidt believes that this type of situation (this "bewilderment" at "feeling" that something is wrong but not being able to justify why) is characteristic of our moral attitudes and that a theory explaining moral judgement must account for these cases. In his opinion, intuitionist theories are better placed to explain them: «Intuitionism in philosophy» he argues «refers to the view that there are moral truths, and that when people grasp these truths they do so not by a process of ratiocination and reflection, but rather by a process more akin to perception, in which one "just sees without argument that they are and must be true"» (HAIDT 2001, 814). Moral judgements are, then, caused directly by these intuitions and reason plays no causal role here.

The indications Haidt presents for doubting that reason usually plays an important role in the formation of moral judgements are as follows:

1) Firstly, it appeals to the growing consensus among psychologists on the fact that, when subjects are trying to solve a problem (and moral problems would be no different from any others on this point), not only is the reasoning process activated, but, simultaneously, another much faster process is activated that brings a quicker solution. This is the "dual processing model". However, Haidt believes there is scientific evidence that, of the two processes, the intuitive and emotional one sets the tone in problem solving, basically through an automatic assessment of situations and people. For example, with respect to the formation of opinions about other people (including judgements about their character or about the moral correctness of them or their actions), the evidence shows that «People form first impressions at first sight [...], and the impressions that they form from observing a "thin slice" of behavior (as little as 5 s) are almost identical to the impressions they form from much longer and more leisurely observation and deliberation» (HAIDT 2001, 820). These first impressions subsequently determine the rest of our moral opinions about these people.
2) Secondly, some research provides evidence that both processes do not occur simultaneously, but that the role of reason is usually limited to justifying or giving meaning, *a posteriori*, to the intuitively issued opinion. And not only is it a *post hoc* reasoning, but also, in Haidt's opinion, it is not reasoning that is genuinely open to evidence and reasons, but one "motivated" by factors such as the desire to maintain good relations with others or to avoid the anxiety that can be produced by seeing our own worldview threatened. Thus, the reasoning process is «more like a lawyer defending a client than a judge or scientist seeking truth» (HAIDT 2001, 820), in the sense that it is not impartial reasoning. Evidence is sought exclusively in favour of our intuitive beliefs.
3) Also, and thirdly, there are also neuropsychological experiments (for example, those by Damasio) that can be taken as indications that the connection between moral reasoning and moral action is quite weak, and that emotions (fundamentally empathy) are much more effective as motivators of moral behaviour, thus proving Hume, Adam Smith and others right.

From his intuitionism and the residual role of reasoning, Haidt draws three important conclusions:

a) First, that the reasoning process only works objectively (and not merely as *a posteriori* rationalisation) under very specific circumstances: it requires the person to have time and capacity, motivation to be accurate (truth-oriented), to lack an *a priori* judgement to defend, and to have neither an interest in maintaining the relationship nor an interest in maintaining coherence (HAIDT 2001, 822). Haidt does not therefore deny that sometimes reasoning (in the classical sense) can be the cause of our moral judgements («particularly among philosophers» Haidt writes, «one of the few groups that has been found to reason well») (HAIDT 2001, 819). But these are rare circumstances. In «more realistic circumstances, moral reasoning is not left free to search for truth» (HAIDT 2001, 822). Often, we resort to reasoning only with the intention of arousing in others the same intuitions that we have ourselves.
b) Secondly, that «our moral life is plagued by two illusions»: the illusion that our moral judgement is due to our reasoning (the "*wag-the-dog illusion*") and the illusion that our arguments change the opinions of others (the "*wag-the-other-dog's-tail illusion*"): it «is like thinking that forcing a dog's tail to wag by moving it with your hand should make the dog happy») (HAIDT 2001, 823).
c) Finally, the lesson Haidt draws is that the roots of human intelligence lie not in our rational ability to search for and evaluate evidence, but «in what the mind does best: perception, intuition, and other mental operations that are quick, effortless, and generally quite accurate» (HAIDT 2001, 822).

But what does intuition really consist of? Drawing on the work of various authors, Haidt proposes the following traits to distinguish between the reflective process and the intuitive process (HAIDT 2001, 818):

The process of intuition	*The process of reflection*
Fast and effortless	Slow and with effort
Unintentional and works automatically	Intentional and controllable
Inaccessible: you are aware only of the result	Accessible to the consciousness and visible
No attention required	Requires attention, which is always limited
Parallel distribution	Sequential
Pattern matching: thinking is metaphorical, holistic	Symbolic manipulation: thought tries to preserve the truth; it is analytical
Common to all mammals	Typical of humans over two years of age and of some apes trained in a certain way
Depends on the context	Does not depend on the context
Depends on the bases (brain and body) that support it	Independent of its bases (can be implemented in any organism or machine acting according to certain rules)

Haidt is also concerned with the mechanisms by which intuitions are "fixed" and with their origin and development. Regarding the former, he draws on Damasio's hypothesis of somatic markers and Lakoff and Johnson's explanation of the role of metaphors in the way we interpret the world.

As is known, according to Antonio DAMASIO (1994) experiences of the world are associated with emotional sensations of pleasure or pain, so they are "marked" as positive or negative. This allows a quick (and automatic) assessment of analogous situations that may arise in the future. There comes a point where the mere thought of a particular action is sufficient to provoke an "as if" response in the brain, whereby the person weakly experiences the same bodily feelings that he or she would experience if performing the action. This "marking" of situations as positive or negative allows the brain to very quickly rule out possibilities for action associated with negative sensations and select those associated with positive sensations much more quickly than rational analysis can do.

Meanwhile, the linguist George Lakoff and the philosopher Mark Johnson have argued that human thought processes are largely metaphorical (LAKOFF & JOHNSON 2003). Metaphors construct a framework of concepts that largely determine our way of interpreting and living in the world. For example, "argument is war" ("Your claims are *indefensible*", "*He attacked every weak point* in my argument", "His criticisms were *right on target*", "*I demolished* his argument", "I've never *won* an argument with him", "If you use that *strategy*, *he'll wipe* you out") (LAKOFF & JOHNSON 2003, 12 f.); "time is money" ("You're *wasting* my time", "This gadget *will save* you hours", "I don't *have* the time to *give* you", "How do you *spend* your time these days?", "I've *invested* a lot of time in her", "I don't *have enough* time *to spare* for that", "He's living on *borrowed* time") (LAKOFF & JOHNSON 2003, 15 f.); "rational is up" while "emotional is down" ("We put our *feelings* aside and had a *high-level intellectual* discussion"; "He couldn't *rise above* his emotions") (LAKOFF & JOHNSON 2003, 24). Many of these metaphorical associations are built from our experiences as physical bodies: for example, we associate the purity and cleanliness we need in food with moral goodness, and corruption and filth with vice HAIDT 2001, 825). In short, what our author seems to want to suggest is that these associations (which Lakoff and Johnson seem to consider automatic and which, in turn, determine our judgements) are behind many of our moral intuitions:

> «Moral intuition, then, appears to be the automatic output of an underlying, largely unconscious set of interlinked moral concepts. These concepts may have some innate basis [...], which is then built up largely by metaphorical extensions from physical experience. Metaphors have entailments, and much of moral argument and persuasion involves trying to get the other person to apply the right metaphor. If Saddam Hussein is Hitler, it follows that he must be stopped. But if Iraq is Vietnam, it follows that the United States should not become involved [...]. Such arguments are indeed a form of reasoning, but they are reasons designed to trigger intuitions in the listener» (HAIDT 2001, 825).

As well as the mechanism for fixing intuitions, Haidt proposes a hypothesis about the evolutionary origin of many of our moral intuitions. His proposal tries to explain them, as Darwin did, as a development of the social instincts of animals. The process would go from mere regularities ("descriptive rules", he calls them) of many species to the "prescriptive rules" (rules for which compliance is reinforced by the reactions of others) that appear in primates and which, with the appearance of language, reach a higher stage (although «the cognitive and emotional machinery of norm creation and norm enforcement was available long before language existed», HAIDT 2001, 826). These are rules related to reciprocity, empathy, altruism, loyalty and so on, and, given the parallelism «between the social lives of humans and chimpanzees, the burden of proof must fall on those who want to argue for discontinuity—that is, that human morality arose ex nihilo when we developed the ability to speak and reason» (HAIDT 2001, 826). However, society carries out "pruning" work on these innate intuitions evolution has inscribed in human beings, modelling them with the influence of others,

immersion in customs and socialisation processes (this explains why children end up with a morality specific to their group or community). Surprisingly, Haidt states that «a culture that emphasized all of the moral intuitions that the human mind is prepared to experience would risk paralysis as every action triggered multiple conflicting intuitions» (HAIDT 2001, 827) (surprisingly, because then evolution does not seem to be wise or adequate enough to ensure survival by moral guidelines alone).

Haidt even proposes six innate intuitions (and their opposites) as the axes of an "intuitive ethics" (which would explain the coincidence of moral rules in different cultures): 1) care (harm); 2) fairness (cheating); 3) loyalty (betrayal); 4) authority (subversion); 5) sanctity (degradation) and 6) liberty (oppression) (HAIDT 2013, ch. 7). Different combinations in the importance assigned to each of these intuitions or emotions would give rise to the different moral systems (HAIDT & CRAIG 2004).

3. *The defeasibility of rules and the intuitionist model*

Legal reasoning is a case of general reasoning with certain particular features. One of these is the importance of what we can call its defeasible character. Defeasibility is not an exclusive aspect of legal reasoning (nor, probably, of normative reasoning), but it can perhaps be considered to have a special importance in the latter. Be that as it may, it is at the centre of a good part of the discussions of current legal theorists[1].

I will call defeasibility the fact that the application of legal rules is necessarily open to unforeseen exceptions, if it is to be rational (fair, reasonable, justified). The norms (rules) are drafted with a general scope; to mark the limits of their field of application, a series of generic properties are taken into account (such as, for example, being over 18 years of age, having acted negligently, having caused damage, having an income greater than x, etc.). These properties (the conditions of application of the rule, to use von Wright's terminology) make it possible to establish the cases in which it is *prima facie* justified to apply the rule and in which cases it is not. In other words, the scope of application of the rule, including its explicit exceptions. However, as it is impossible for legislators to foresee all the properties that will be relevant in order to satisfactorily regulate a case, it may be that the application of the rule to a specific case (to which it should initially be applied) is not reasonable, that is, it appears to be unjustified as a relevant property that had not been considered is present in that case. Obviously, in the case of law, the reasoning that shows the application of the rule to that particular case is not justified must, in turn, be based on other legal standards (values, principles, reasons underlying the rule, etc.). Thus, as well as the explicit exceptions provided for in the formulation of the rule, other exceptions not provided for appear after a more in-depth analysis of the case to be resolved. Following a terminology proposed by Celano, we can therefore distinguish between "normal" and "non-normal" cases of application of the rule (in the latter the rule would be displaced). But what criteria do we use to distinguish between the two, and do the cognitive sciences have anything to contribute to this problem?

In my opinion, projecting the intuitionist model to the problem of defeasibility is a good testing ground for assessing its role in the analysis of legal reasoning. In the philosophy of law, in several works Bruno Celano and Marco Brigaglia have suggested a "psychologized" theory of defeasibility, in some points close to intuitionism (although Haidt's intuitionism goes further).

> «The hypothesis is the following: the fact that certain cases are normal and others are not—on this, as has been said, depends the answer to the question whether it is reasonable, in such a case, to

[1] See, for example, the contributions in FERRER & RATTI 2012.

> reconsider the rule (this is, I emphasize again, a normative question)—depends on psychological facts: certain cases are normal or abnormal only because they appear to us as such. Which cases appear to us as normal and which do not is a clear question of psychology. So: whether or not a certain rule, in a given case, is a reason to act – a justifying reason; this is the crucial point – depends on our psychological make up» (CELANO 2017, 101, my translation; see also CELANO & BRIGAGLIA 2017).

For these authors, mental states can be conscious and unconscious. What allows us to detect that a case is not normal is a feeling of "surprise" or "bewilderment" (to use Haidt's expression), which is the conscious manifestation of certain unconscious mental processes.

To analyse this type of proposal, it may be useful, once again, to distinguish their descriptive, conceptual and normative scope:

a) *Descriptive level:* From the intuitionist model it is possible to account for the ability to distinguish between normal and non-normal cases of application of a rule based on the notion of intuition and emotion. It is intuition that alerts us that something is wrong with the application of the norm and we "discover" or "realise" that a case is or is not normal through a "feeling" of approval or disapproval of the possibility of the rule being applied to that case. According to Haidt's intuitionism (going beyond that suggested by Celano and Brigaglia), the intuitionist model could further postulate that such intuition would be part of (or might be traced back to) a set of moral principles that have an evolutionary origin, subsequently shaped by social and cultural factors (especially in the case of law, given its institutional character). What these intuitions and emotions would be telling us is that the application of the rule to that case would violate some relevant moral and/or legal principle which would ultimately be linked to the survival of the species. On the other hand, legal reasoning requires that the "abnormal" case (as well as the proposed solution) should be justified on the basis of legal standards (i.e., principles or values recognised in the legal system), and not on the basis of feelings. But, if Haidt is right, this justification will be an *a posteriori* rationalisation that does not change the direction of the initial intuition.
b) *Conceptual level*: Therefore, we can call "normal cases" those in which the application of the solution established in the rule generates a feeling of approval, and non-normal cases (cases in which there are implicit exceptions) those in which the application of the solution established in the rule generates a feeling of disapproval. The feeling (or intuition) could be taken not only as a warning (a symptom) that we are facing a case with an implicit exception, but also as a conceptual criterion for identifying such a case.
c) *Normative level:* At this level it would be argued that it is justified to displace (defeat) the rule when its application to a particular case generates a feeling of disapproval (on the other hand, it would be justified to apply the rule when it does not generate a feeling of disapproval). As can be seen, this step implies a reduction of the normative level—the justification level—to the psychological level.

Is this an adequate reconstruction of the defeasible nature of legal reasoning? I believe it inherits some general problems from the intuitionist model.

4. *Some doubts about the plausibility of the intuitionist model: descriptive claims*

As we have seen, the intuitionist model "à la Haidt" can be characterised using four theses:

1) *Intuitionism:* our moral judgment is dominated, or strongly influenced, or largely conditioned, by quick, unconscious, unintentional intuitions, a long way from reflection.

2) *Emotivism:* these moral intuitions depend fundamentally on our emotions or are the expression of these emotions.
3) *Innatism:* these are innate intuitions and emotions, transmitted genetically, although they can be culturally moulded.
4) *Darwinism:* moral intuitions are seen as mechanisms that evolution has selected because they ensure the survival of the species.

Haidt's explanation of how we reason in normative contexts is based on these theses. However, in my opinion, all of them raise certain problems, which could cast doubt on whether the intuitionist model is a good description or explanation of our reasoning processes in these contexts. Several of these objections arise from a philosophical analysis of the data used by the cognitive sciences to draw their conclusions (which shows, by the way, the need for an interaction between both viewpoints). Let us take a look at them:[2]

1) *Intuition and reason:* Haidt's first thesis seems to establish too radical a dichotomy between intuition and reason. In other words, he seems to be assuming that one excludes the other, as if they were two independent information processing systems, in contrast to the views of other authors, such as Kahneman and Damasio. Haidt's intuitionist model also adds that decisions are made in System 1, while System 2 provides only an *a posteriori* rationalisation of decisions. However, it also seems a plausible thesis that the two systems collaborate in decision making, either because System 2 can correct System 1 (Kahneman) or because emotions simplify and order information and alternative actions based on past experiences, avoiding complexity that would paralyse reason (Damasio). This hypothesis of collaboration between the two systems fits well into philosophical analyses of intuition such as that of Mario Bunge, who has shown that the idea of "intuition" encompasses many different phenomena (modes of perception; forms of imagination; sudden, rapid or incomplete inferences; capacity for synthesis; capacity to evaluate a situation and to choose the best alternatives, and so on) (BUNGE 2013, ch. III.1) and several of these forms of intuition cannot be seen as opposing or excluding reason: rapid or incomplete inference is embryonic or primitive reasoning and "synoptic apprehension", as Bunge points out, «is not a substitute for analysis, but a reward for careful analysis». But, in addition, the different types of intuition—even those that cannot be seen strictly speaking as reasoning, not even incomplete reasoning—are especially encouraged by the continuous exercise of reasoning, problem analysis, experience in an activity, or dedication to studying a discipline. In short, many intuitions would not be possible without collaboration between the two systems.
2) *Normative intuitions and emotions:* Some difficulties also arise over the relationship between normative judgements and emotions. The argument behind this linkage is as follows: On one hand, brain imaging shows that when we reason morally, areas of the brain related to emotion are strongly activated; on the other, psychological tests show that in most cases we solve moral dilemmas intuitively. The two things must be related: intuitions, therefore, arise from emotions—from the activity of the affective areas of the brain. But in my opinion this way of reconstructing moral decisions raises several problems:

a) First, the evidence we have for the role of emotions in decision-making is actually indirect. What neuroscientists can prove is that, when we are faced with a moral dilemma, the areas of the brain related to emotions are activated in a particularly pronounced way. But this does not allow us to infer that emotions generate moral judgment. It could be that it is the moral

[2] A development of these criticisms can be found in GONZÁLEZ LAGIER 2017.

judgment that generates the emotion: for example, realising, through analysis, the unfairness of a situation may generate indignation; or, in a moral dilemma, being aware that any solution will cause some sort of harm may provoke regret.

b) Secondly, the notion of emotion apparently assumed by the intuitionist model converts emotions into mere feelings (of pleasure or displeasure) that lack propositional content. This way of understanding emotions and their relation to morality departs from the conceptions of emotion most widespread today among philosophers and many psychologists. For these conceptions (which, to a large extent, support Aristotle's conception of emotions), at least three distinct dimensions must be distinguished: (i) A cognitive dimension (in a broad sense, ranging from a belief to a mere perception); (ii) an affective or purely phenomenological dimension (the sensation of pleasure or pain) and (iii) a motivational dimension (a tendency towards action). Cognitive theories of emotion focus on the first aspect, "somatic" and mechanistic theories of emotion focus on the second, and behaviourist theories on the third element. Finally, non-reductivist theories try to account for all aspects of emotions. If emotions are identified exclusively with the second aspect, many problems remain to be solved: (a) the possibility of unconscious or sensationless emotions is not accounted for; (b) the possibility that emotions can be part of rational (teleological) and not only causal explanations of behaviour is not accounted for; (c) the two-way nature of the relationship between emotions and beliefs is not accounted for; (d) the possibility of evaluating emotions as reasonable or unreasonable (depending on the underlying belief) is not accounted for; and, above all, (e) there is, once again, a marked distancing between emotions and reason (González Lagier 2009, ch. II).

c) *Innatism:* Nor is the thesis—closely linked to the previous ones—of the innatism of moral intuitions free from objections. The arguments in favour of innatism again rest on the automatic nature of the response and the inability to give reasons (Haidt), together with the coincidence of responses despite the diversity of the respondents (Hauser). However, to be sure that our moral beliefs are innate, we would have to entirely rule out the possibility that the automatic responses are due to the acceptance of deeply rooted but culturally transmitted principles. As Adela Cortina, commenting on Haidt's experiments, points out:

> «The fact that respondents answer intuitively—that is, immediately and automatically, without being aware of how they have come to formulate the judgment—and that on many occasions they can give no reason why an action seems good or bad to them, may well be explained by the fact that they have learned it socially and have not subjected it to review» (Cortina 2006, 86, my translation).

Many rules of social morality (such as incest, in Haidt's example) are inherited from the social environment, which, most of the time, inculcates them without giving reasons to justify them. On the other hand, the universality of rules or principles must also be taken with caution: many supposedly universal moral behaviours are not, in fact, moral ones, or their universality is independent of their moral nature. For example, Marc Hauser has postulated the universality of the offspring care principle (which obviously has a clear evolutionary explanation). But should we also draw the conclusion that foraging or fleeing from predatory animals also has moral value (Barta 2013, ch. III)? Or reproduction? What I mean is that the nature of many habits given as examples of universally accepted behaviours is not (or not exclusively) moral.

d) *Darwinism:* We have seen that another of the characteristic theses of this attempt to account for morality from a biological and neuroscientific point of view is, in fact, moral Darwinism. Moral intuitions are seen as mechanisms that evolution has selected because they ensure the survival of the species. This thesis, however, suffers from a degree of ambiguity. To clarify it, this is useful to distinguish, once again, between the (descriptive) claim to explain

> morality and the (normative) claim to justify it. We can call the former descriptive moral Darwinism and the latter prescriptive moral Darwinism. Descriptive moral Darwinism, in turn, can try to explain human beings' *capacity* to demonstrate ethical behaviour (to evaluate behaviours as right or wrong from a moral point of view) or (more ambitiously) to try to explain *the content* of morality; in other words, why we believe certain behaviours are right or why some principles or values are so widespread (AYALA 2013, 61). Descriptive moral Darwinism appeals to the idea that behaving morally or adapting one's behaviour to certain principles is a trait that has facilitated the evolution of the species and its survival. The prescriptive version adds that, as this is so, such principles are justified (we will return to the prescriptive issue in the next section).

Descriptive moral Darwinism is merely a hypothesis which has not been sufficiently confirmed. Take the claim that what explains the human ability to evaluate behaviours as good or bad and to adjust behaviour to certain principles is that this ability is an evolutionary advantage. To accept it conclusively, one would have to reject the (also plausible) alternative hypothesis put forward by Francisco Ayala according to which ethical behaviour is only indirectly a result of evolution in as far as it is a consequence of the development of human intelligence; in other words, what has adaptive value and has been favoured by evolution is human intelligence, not the ability to behave morally (which is a consequence of human intelligence) (AYALA 2013, 66). If, on the contrary, the claim is that moral codes are determined by evolution, the problem is that it is not possible to find a set of relevant principles that are universal, but not formulated in an excessively vague and empty way. Moreover, it is possible to find types of behaviour, such as aggressiveness and territoriality, that are evolutionarily important and cannot be accepted as examples of moral behaviour.

In short, I believe that intuitionism, at least this version of it, is more of a hypothesis still lacking confirmation and in need of refinement than a well-founded explanation of the functioning and basis of practical reasoning.

5. *The problem of justification and normativity*

Studies of practical reasoning carried out in the cognitive sciences can adopt three positions regarding the ideas of correctness or justification of reasoning—in other words questions of normativity: (1) limiting themselves to trying to give a description and explanation and giving up their normative claims; (2) trying to draw conclusions about justification from those descriptions or (3) denying that it makes sense to speak of ideas such as justification, correctness or normativity. Michel Ruse has explained how cognitive sciences can lead to this third option:

> «Is it not the case that sometimes, when one has given a causal explanation of certain beliefs, one can see that the beliefs, in themselves, neither have a foundation nor could ever have such a foundation? [...] Once we see that our moral beliefs are simply an adaptation put in place by natural selection, in order to further our reproductive ends, that is an end to it. Morality is no more than a collective illusion fobbed off on us by our genes for reproductive ends» (RUSE 1991, 506).

It seems to me that what lies behind these positions that challenge the idea of justification and normativity is the following: if practical reasoning is *necessarily* determined by intuitions, heuristics and unconscious and uncontrollable biases—that is, if it cannot be otherwise—then there is no point in asking when it is correct or under what conditions it is justified. It is possible to speak of regularities and divergences (which would have to be explained causally), but not of norms, rules or criteria for correctness. In short, this is a sceptical stance on

normativity. Here I will set aside this option, focusing on the problems with the second alternative: the claim that the cognitive sciences can have a strong normative scope.

5.1. *Normative claims based on the cognitive sciences*

Can the cognitive sciences say anything about when reasoning is correct and when it is not? Haidt seems to have only descriptive pretensions, but other authors have gone further and have tried to find in moral intuitions—or directly in the functioning of the brain—normative criteria with which to evaluate the correctness of our moral theories or principles about how we should live. Some examples include:

Neuroscientist William Casebeer, who has claimed that the Aristotelian moral theory of virtue is more plausible from a neurobiological point of view than the moral theories of Kant or John Stuart Mill. His argument is that each of these theories implicitly contains a specific moral psychology that demands different cognitive capacities. Thus, Kant's theory would seem to require «at least the ability to check universalized maxims for logical consistency in a manner that is separable from the taint of affect and emotion»—an ability that corresponds to the functions of the frontal region of the brain. Mill's utilitarian theory requires the ability to perform utilitarian calculations and cultivate emotions that move us to procure the happiness of others, which involves the pre-frontal, limbic and sensory regions of the brain. The Aristotelian ethics of virtue, finally, would be the most demanding, because it requires us to educate our character so that our appetites are in line with good reasons. This implies a "global psychology" that requires the coordinated intervention of the aforementioned regions of the brain. Our author believes there is scientific evidence to tentatively accept that "moral cognition" brings different brain systems and networks related to both cognition and emotions into play in a coordinated way (i.e., the pre-frontal, frontal, limbic and sensory regions: what could be called "the area of moral cognition"), showing that «there is clear consilience between contemporary neuroethics and Aristotelian moral psychology» (CASEBEER 2003, 845). This makes it possible to rule out the plausibility of the other theories.

For his part, Michel Gazzaniga writes: «I would like to support the idea that there could be a universal set of biological responses to moral dilemmas, a sort of ethics built into our brains. My hope is that we soon may be able to uncover those ethics, identify them, and begin to live more fully by them» (GAZZANIGA 2005, xix).

Meanwhile, Patricia Churchland suggests that, just as health «is a domain where science can teach us, and has already taught us, a great deal about what we ought to do», so too in «the domain of social behavior [...] we may learn a great deal from common observation and from science about conditions favoring social harmony and stability, and about individual quality of life» (CHURCHLAND 2011, 190).

5.2. *Is it possible to infer normative conclusions from purely descriptive statements? Arguments against Hume's Law*

I will now make some comments on the normative scope of the cognitive sciences.

As is well known, in a famous passage, Hume states that:

> «In every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surprized to find, that instead of the usual copulations of propositions, *is*, and *is not*, I meet with no proposition that is not connected with an *ought*, or an *ought not*. This change is imperceptible; but is, however, of the last consequence. For as this *ought*, or *ought not*, expresses some new relation or affirmation, it is necessary

> that it should be observed and explained; and at the same time that a reason should be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it» (HUME 1960 [1739-40], 469).

This passage has usually (although not unanimously) been interpreted as prohibiting the inference of statements "about what ought to be" from statements "about what is the case". Thus, some of the authors who defend the normative relevance of the cognitive sciences try to argue against the validity of Hume's Law. Let us examine in more detail these arguments, which are usually of two types:

1) The first group involves *arguments based on counterexamples*: A frequent way of showing that it is possible to ground rules or values in descriptions is to present examples of arguments in which this derivation is apparently made. This is the strategy followed, in a famous article, by John Searle (SEARLE 1964). Cognitive scientists have also resorted to this type of argument. Marc Hauser, proposes the following:

> «FACT: The only difference between a doctor giving a child anesthesia and not giving her anesthesia is that without it, the child be in agony during surgery. The anesthesia will have no ill effects on this child, but will cause her to temporarily lose consciousness and sensitivity to pain. She will then awaken from the surgery with no ill consequences, and in better health thanks to the doctor's work.
> EVALUATIVE JUDGMENT: Therefore, the doctor should give the child anesthesia» (HAUSER 2006, 3)

However, this type of argument appears to fall into one of the following errors: (1) it confuses what is good or owed from a technical point of view with what is good or owed from a moral or normative point of view; (2) it presents as a complete argument what is, in fact, an incomplete argument that includes a hidden premise: the very norm or value judgement from which the conclusion is derived.

To avoid the first error, it should be noted that every time a statement includes the term "must" it is not necessarily a genuinely normative statement. Sometimes "must" expresses a conjecture ("must be so" can mean "probably is so"). At other times, it can be replaced by "has to" and expresses a practical necessity. It is important to distinguish between *deontic* or genuine *duties* and *technical*, prudential *duties* or *practical necessities*. Many of the examples given as derivation of "must" from "is" do not conclude genuine deontic duties, but rather practical necessities. As von Wright points out, we can

> «find two main answers to the question of why a certain thing should or may or may not have to be done. One is that there is a rule ordering or permitting or prohibiting the doing of the thing. The other is to say that the ends and necessary connections make the doing or not doing of the thing a practical necessity (or not)» (VON WRIGHT 1963, 74).

Regarding the second error, it is easy to see that the arguments we formulate in everyday contexts often do not include all their premises. It is even feasible to think that in some cases it is impossible in practice to state all the premises necessary to reach the conclusion. But what follows from this is the *defeasibility* or revisability of the conclusion, not that its correctness is independent of the implicit premise. In order to be correct, Hauser's example presupposes a normative premise that unnecessary suffering should be avoided.

2) The second type of argument *restricts Hume's Law to deductive arguments*: it is claimed that what Hume's Law proscribes is deductively deriving a duty from existence, but there are other types of acceptable inferences, such as induction or inference of the best explanation,

> by means of which it is possible to move from descriptions of facts to rules (CASEBEER 2003, 842 f.; CHURCHLAND 2011, ch. 1). One way of arguing that Hume's Law refers exclusively to deductive inferences is to see it as a consequence of the logic conservation principle: in a deduction it is not possible to conclude something that was not already included in the premises. Deductions can make us aware of new facts, but they must already be in the premises. We cannot therefore deduce ought-to-be statements from descriptive propositions alone. This can happen with anything. As Pigden observes «You can't get 'hedgehog' conclusions from hedgehog-free premises (at least, not by logic alone)» (PIDGEN 1991, 423). However, unlike deductions, inductions and inferences to the best explanation do extend our knowledge, so the principle of conservation does not apply to them. Therefore, if Hume's Law is only a manifestation of the principle of conservation of deductive arguments, then those who argue that it is not applicable to non-deductive inferences are right.

But is that all Hume's Law is? Probably not. It can be argued that there are important differences between descriptive statements and normative statements—sometimes we speak of a "logical gulf" between them, or between facts, on the one hand, and rules and values, on the other. Thus, descriptive statements have a downward direction of fit[3] (i.e., words to world: words are intended to fit the world) while normative statements have an upward direction of fit (i.e., world to words: the world is intended to fit the words). Descriptive statements are true or false, while rules or values are not. Statements expressing duties presuppose an internal point of view, in which deontic terms (obligatory, forbidden) are used, whereas if a description refers to a rule or a duty, it does so from an external point of view in which deontic terms are merely mentioned. All these differences between descriptions and rules mean that the former cannot serve as reasons, or express reasons, to justify the latter. It is not only that deductive justification requires that what is to be deduced is included among the premises, it is that, even when it is admitted that induction or abduction can have a justificatory scope, no descriptive statement can, by itself, be a reason justifying a prescriptive statement. It can provide an explanatory reason why we accept certain rules or values. But if one wants to conclude the justification of a rule from descriptions and by means of non-deductive arguments, one has the burden of proof.

If the above considerations are correct, when it is proposed that we should follow patterns of behaviour that have adaptive value, either these are simply recommended as prudent measures to maintain the survival of the human species but without any genuinely normative character, or it is assumed that the survival of the species is a morally valuable end, in which case the normativity stems not from the facts, but from this value assumption.

[3] An explanation of the idea of "direction of fit" can be found in SEARLE 1975.

References

ATIENZA M. 2017. *Epílogo (abierto)*, in AGUILÓ REGLA J., PEDRO GRÁNDEZ P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, 429 ff.

AYALA F. 2013. *Evolución, ética y religión*, Universidad de Deusto.

BARTA R. 2013. *Cerebro y libertad. Ensayo sobre la moral, el juego y el determinismo*, Fondo de Cultura Económica.

BUNGE M. 2013. *Intuición y razón*, Debolsillo.

CASEBEER W.D. 2003. *Moral Cognition and Its Neural Constituents*, in «Nature Reviews Neuroscience», 4, 840 ss.

CELANO B. 2017. *Particularismo, psicodeóntica. A propósito de la teoría de la justificación judicial de Manuel Atienza*, in AGUILÓ REGLA J., PEDRO GRÁNDEZ P. (eds.), *Sobre el razonamiento judicial. Una discusión con Manuel Atienza*, Palestra, 59 ss.

CELANO B., BRIGAGLIA M. 2017. *Reasons, rules, exceptions: toward a psychological account*, in «Analisi e diritto 2017», 131 ff.

CHURCHLAND P. 2011. *Braintrust*, Princeton University Press.

CORTINA A. 2011. *Neuroética y neuropolítica. Sugerencias para la educación moral*, Tecnos, 2011.

DAMASIO A.R. 1994 *Descartes' Error: Emotion, Reason and the Human Brain*, Avon Books.

FARIÑA F., ARCE R., NOVO M. 2002. *Heurístico de anclaje en las decisiones judiciales*, in «Psicothema», 14, 1, 39 ff.

FERRER J., RATTI G. (eds.) 2012. *The Logic of Legal Requirement. Essays on Defeasibility*, Oxford University Press.

GARCÍA CAMPOS J. 2009. *Justificación y racionalidad desde la teoría dual del conocimiento,* «Ideas y valores», 139, 61 ff.

GARCÍA CAMPOS J. 2011. *Razonamiento y racionalidad desde la psicología evolucionista*, in «Metatheoria», 2, 1, 79 ff.

GARCÍA CAMPOS J. 2012. *Convergencias y divergencias en las teorías duales de sistemas*, in «Andamios», 9, 19, 283 ff.

GARCÍA CAMPOS J. 2014. *Normatividad y psicología cognitiva. La propuesta naturalizada de M. Bishop y J. D. Trout*, in «Revista de Filosofía de la Universidad de Costa Rica», 53, 135, 25 ff.

GAZZANIGA M.S. 2005. *The Ethical Brain*, Dana Press.

GONZÁLEZ LAGIER D. 2009. *Emociones, responsabilidad y Derecho*, Marcial Pons.

GONZÁLEZ LAGIER D. 2017. *A la sombra de Hume. Un balance crítico del intento de la neuroética de fundamentar la moral*, Marcial Pons.

HAIDT J. 2001 *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, in «Psychological Review», 108, 4, 814 ff.

HAIDT J., CRAIG J. 2004. *Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues*, in «Daedalus», 133, 4, 55 ff.

HAIDT J. 2013. *The Righteous Mind. Why Good People Are Divided by Politics and Religion*, Penguin.

HAUSER M. 2006. *Moral Minds: The Nature of Right and Wrong*, Harper Collins.

HUME D.1981. *A Treatise of Human Nature*, ed. by L.A. Selby-Bigge, Clarendon Press. (Originally published in 1739-40)

KAHNEMAN D. 2003. *Maps of Bouded Rationality: Psychology for Behavioral Economics*, in «The American Economic Review», 93, 5, 1449 ff.

KAHNEMAN D. 2011. *Thinking, Fast and Slow*, Penguin. (*Pensar rápido, pensar despacio*, in

LAKOFF G., M. JOHNSON, 1980. *Metaphors We Live By*, The University of Chicago Press.

PIGDEN C.R. 1991. *Naturalism*, in SINGER P. (ed.), *A Companion to Ethics*, Blackwell, 421 ff.

RUSE M. 1991. *The Significance of Evolution*, in SINGER P. (ed.), *A Companion to Ethics*, Blackwell, 500 ff.

SEARLE J.R. 1964. *How to Derive an 'Ought' from an 'Is'*, in «Philosophical Review», 73, 1, 43 ff.

SEARLE 1975. *A Taxonomy of Illocutionary Acts*, in GUNDERSON K. (ed.), *Language, Mind and Knowledge*, University of Minnesota Press, 344 ff.

VON WRIGHT G. H. 1963. *The Varieties of Goodness*, Routledge & Kegan Paul.