

Humean Rationality: More Than Purely Instrumental?¹

José Luis TassetProfessor of Moral Philosophy, University of A Coruña, Spain  <https://dx.doi.org/10.5209/ijhe.102540>

Recibido: 30/3/2025 • Revisado: 18/5/2025 • Aceptado: 4/6/2025

EN Abstract. David Hume's theory of practical rationality has been analysed in detail in the history of ethical, political, and economic thought. It is often ascribed a purely instrumental conception of rationality, in which practical reasons are merely instruments to achieve desired ends. This paper aims to challenge this interpretation by arguing that while Hume criticises the role of reason in practical decision-making, his theory of practical rationality is much more complex and contains normative elements.

In his *Lectures on the History of Moral Philosophy*, John Rawls formulates and develops this standard interpretation of Hume in a paradigmatic way by claiming that his theory of rationality is purely empirical, psychological and instrumental. Rawls argues that Hume does not provide a solid foundation for morality, leading to the conclusion that practical rationality is simply a matter of coherence between means and ends. However, it can be argued conclusively that Rawls ignores the complexity of Humean theory itself. Hume's critique of reason does not imply that it lacks a normative foundation. Rather, it suggests a reconfiguration of how we should understand morality and rationality.

Keywords. David Hume; John Rawls; instrumental rationality; normative rationality; scepticism; levels of morality; levels of rationality.

JEL Code: B12; B15; C70.

ES La racionalidad humeana: ¿más que puramente instrumental?

ES Resumen. La teoría de la racionalidad práctica de David Hume ha sido analizada en detalle en la historia del pensamiento ético, político y económico. A menudo se le atribuye a Hume una concepción puramente instrumental de la racionalidad, en la que las razones prácticas son meros instrumentos para alcanzar los fines deseados. Este artículo tiene como objetivo cuestionar esta interpretación, argumentando que, si bien Hume critica el papel de la razón en la toma de decisiones prácticas, su teoría de la racionalidad práctica es mucho más compleja y contiene elementos normativos.

En sus *Lecciones sobre la historia de la filosofía moral*, John Rawls formula y desarrolla esta interpretación estándar de Hume de manera paradigmática al afirmar que su teoría de la racionalidad es puramente empírica, psicológica e instrumental. Rawls sostiene que Hume no proporciona una base sólida para la moralidad, lo que lleva a la conclusión de que la racionalidad práctica es simplemente una cuestión de coherencia entre medios y fines. Sin embargo, se puede argumentar de manera concluyente que Rawls ignora la complejidad de la propia teoría humeana. La crítica de Hume a la razón no implica que esta carezca de una base normativa. Más bien, sugiere una reconfiguración de cómo debemos entender la moralidad y la racionalidad.

Palabras clave. David Hume; John Rawls; racionalidad instrumental; racionalidad normativa; escepticismo; niveles de moralidad; niveles de racionalidad.

Códigos JEL: B12; B15; C70.

¹ I have dealt extensively with human practical rationality in the context of defining a possible utilitarian paradigm of rationality and in a critical engagement with Bentham's conception of Hume's moral and political philosophy in particular. Some of the general ideas in this paper are drawn from previous works, some of which have already been published, but others are still in progress. Cfr. (Tasset, 2011); (Tasset, 2018); (Tasset, 2019); and finally (Tasset, 2022). I would like to thank Alfonso Palacio Vera for his precise comments on a first version of this work.

PT Racionalidade Humana: Mais Do Que Puramente Instrumental?

PT Resumo. A teoria da racionalidade prática de David Hume foi analisada em detalhe na história do pensamento ético, político e econômico. Frequentemente, atribui-se a Hume uma concepção puramente instrumental da racionalidade, na qual as razões práticas são meros instrumentos para alcançar fins desejados. Este artigo tem como objetivo questionar essa interpretação, argumentando que, embora Hume critique o papel da razão na tomada de decisões práticas, sua teoria da racionalidade prática é muito mais complexa e contém elementos normativos.

Em suas *Lições sobre a História da Filosofia Moral*, John Rawls formula e desenvolve essa interpretação padrão de Hume de maneira paradigmática ao afirmar que sua teoria da racionalidade é puramente empírica, psicológica e instrumental. Rawls sustenta que Hume não fornece uma base sólida para a moralidade, o que leva à conclusão de que a racionalidade prática é simplesmente uma questão de coerência entre meios e fins. No entanto, pode-se argumentar de forma conclusiva que Rawls ignora a complexidade da própria teoria humiana. A crítica de Hume à razão não implica que ela careça de uma base normativa. Pelo contrário, sugere uma reconfiguração de como devemos entender a moralidade e a racionalidade.

Palavras-chave: David Hume; John Rawls; racionalidade instrumental; racionalidade normativa; ceticismo; níveis de moralidade; níveis de racionalidade.

JEL classificação: B12; B15; C70.

Sumário: 1. Introduction; 2. The origin of the attribution of an instrumental theory of reason to Hume: the radically sceptical interpretation of his philosophy; 3. Outline and discussion of a hypothetical theory of practical rationality; 4. A hypothetical normative development of a post-Humean nature; 5. Conclusions.

Cómo citar: Tasset, J. L. (2025): Humean Rationality: More Than Purely Instrumental? *Iberian Journal of the History of Economic Thought*, 12(1), 7-17. <https://dx.doi.org/10.5209/ijhe.102540>

1. Introduction

Hume's theory of practical rationality has had and continues to have a major influence on current debates in moral philosophy, philosophy of mind and the social sciences in general, especially in recent studies of economic rationality. In its standard form as a critique of classical moral rationalism, it has served to challenge the simple isomorphism between morality and rationality (moral claims would be claims of practical reason itself) and pure moral cognitivism (moral judgements can be strictly true or false). His general questioning of the role of reason in practice, from its initial formulation in the *Treatise of Human Nature* (1739-40) to its mature form in the *Inquiry Concerning the Principles of Morals* (1751), is part of any minimally defensible formulation of the modern theory of rational choice. Even in the field of the history of economic thought, it is often claimed that the so-called "standard theory" has its roots in the hypothetical Humean theory of instrumental rationality.

But the very active role of Hume's theory of practical rationality in the contemporary discussion of

"agency" has a price. Very often and in supposed opposition to an Aristotelian or Kantian theory of practical rationality (as a paradigmatic example, let us contrast the Humean theory with that formulated by John Rawls in his *Lectures on the History of Moral Philosophy*), Humean theory is considered to defend an essentially instrumental or hypothetical conception of practical reasons, the core of which is the idea that the only thing required of an agent in order to be rational is a coherent form of means-ends fit, and that no end as such is rationally justified in a negative or positive sense.²

In general, I think it can be shown that Hume, as a critic of the role of reason in practical decision-making, whether moral, political and/or economic, has a theory of practical rationality and that this has a substantive and normative character,³ not just an instrumental one, albeit in complex ways and perhaps not as directly as some would like or wish. But what one might ask of Hume's philosophy for today is another question that we cannot and do not wish to address here.

² Even without considering the internal development of Hume's theory from his main work (1739-1740) to the formulations of the 1750s, which I do not think should be done, it is not easy to determine whether (a) Hume was simply a radical sceptic of practical (and/or theoretical) reason, even instrumental reason; or whether (b) Hume finally accepted and applied a distinction between the "reasonable" and the "unreasonable" that went beyond the rational/irrational distinction as a mere means/end adaptation, and thus proposed a theory of practical rationality of a substantive and not merely instrumental character. It should be added that his critique of reason, at least in practice, tends to question the scope of this rationality and not necessarily its existence or its possibilities.

³ I am not contrasting "substantial" and "procedural" here, as Herbert Simon does, for example. By "substantial" and "normative" rationality I mean the opposite of "instrumental", i.e. in the context of this work a theory that deals with the ends of action and not just the means.

What we will discuss in this short paper is the following: We will question the attribution to Hume of a pure instrumentality in the understanding of practical reason, a charge paradigmatically formulated by John Rawls, on whose brilliant expositions we will mainly focus, and secondly, we will outline in broad strokes what we understand as a Humean theory of practical and especially moral rationality, which is normative and not purely instrumental in nature and mainly revolves around the theory of the “judicious spectator” or “impartial observer”, one of the most important conceptual insights of Hume’s moral and political philosophy and of modern practical thought.⁴

Finally, we will even sketch a possible development of an important normative implication for practical rationality that derives directly from Hume’s arguments, although it is, in a sense, beyond what he explicitly mentions.

2. The origin of the attribution of an instrumental theory of reason to Hume: the radically sceptical interpretation of his philosophy.

John Rawls’ interpretation of David Hume’s epistemology is characterized by a shift from initial admiration⁵ to a deep critique that suggests Hume’s philosophical stance is both banal and aporetic. Rawls acknowledges the interpretative challenges surrounding Hume, noting that scholars have approached his work from various angles. He aligns himself with the classical sceptical interpretation while integrating Norman Kemp Smith’s naturalistic perspective, which emphasizes Hume’s psychological naturalism over his scepticism.⁶

The radically sceptical interpretation of Hume’s epistemology, which John Rawls takes as his starting point and which was the most traditional until the 20th century, touches on the fact that Hume would deny the possibility of an authentically rational knowledge of nature and reality, even of the external world. Our apparent knowledge of this reality would in fact be based on structures that are not strictly rational, above all on memory and habit. We find the model of this skeptical critique of knowledge in Hume’s anal-

ysis of induction and causality; both would be based on the mere habit of expecting events to repeat themselves regularly, and in both cases absolute certainty would be ruled out. This very restricted view of rationality would extend in a second stage to our practical rationality, in which Hume would apparently deny the possibility that moral as well as aesthetic or political judgements can be based on anything other than our feelings, emotions or passions. As Hume argued in the justly famous passage on the slavery of reason at the *Treatise*, our judgements in epistemology as well as in ethics and politics, and thus also in economics, which was part of moral philosophy in Hume’s time, would be a projection of our feelings rather than a result of any kind of reason or rationality.

This skeptical, nuance-free version was joined in the 20th century by the idea that what reason is incapable of achieving should be placed in the hands of our common human nature. The nature of human beings, their deep and enduring psychological structures, for example moral feelings such as sympathy, benevolence or, again, habit, would save us from sinking into radical skepticism and distance us from doubt in the practical realm. Skepticism and naturalism would thus be two sides of the same phenomenon: the impossibility of regulating our lives through reason.

But, in reality, Hume wanted to say something quite different and much simpler. Against Descartes and the so-called dogmatic rationalists (mainly Samuel Clarke), Hume claimed that rationality, certainty, especially mathematical certainty, is only a very small part of our rationality and not even the dominant one. In the face of this complete certainty, reason opens up the immense realm of probable reason, of probability, both in the epistemological and in the practical spheres. And Hume thinks that this is enough to avoid falling into either skepticism or dogmatism. To simplify the common humean argument: Moderate skeptical criticism (awareness of the limits of reason) and empiricism (primacy of sensory experience) move us away from certainty without allowing us to lapse into irrationality.⁷

⁴ The ideal observer theory (IOT) offers an explanation for the truth/objectivity of moral and value judgements in terms of the approval or disapproval of an ideal observer. It is common to assume that this theory is explicitly addressed by Adam Smith and Henry Sidgwick. John Rawls, however, believes that the first theorist of impartiality and its role in moral judgments is David Hume. Hume does not use the term “impartial observer”, but rather “judicious spectator”, in Book III of the *Treatise*, i.e. 1740; Adam Smith uses the term “impartial observer” explicitly, albeit much later, in the *Theory of Moral Sentiments* of 1759, 2nd edition of 1761 and 6th edition of 1790.

⁵ “Hume says that the work was projected before he left the university at fifteen in 1726, it was planned before he was twenty-one in 1732, and it was composed before he was twenty-five in 1736. These dates are probably not entirely accurate but are extraordinary nonetheless. (...) These astounding facts leave one speechless.” (Rawls, 2000, p. 21)

⁶ In relation to Hume’s overall view of philosophy, Rawls concedes that “Hume has not been easy to interpret.” (Rawls, 2000, p. 21). He then explains that his interpretation follows the line of Hume’s classical skeptical interpretation, which was later supplemented by Kemp Smith’s naturalistic version (Smith, 2005), which is a further development of the skeptical view.

For Rawls, Kemp Smith “gave pride of place to Hume’s psychological naturalism and de-emphasized his skepticism.” (Rawls, 2000, p. 22). Among the authors after Kemp Smith, he also cites Burnyeat and Fogelin, who “emphasizes both skepticism and naturalism, and indeed sees them as complementary and working together.” (Burnyeat, 1983; Fogelin, 2009; Rawls, 2000, p. 22) A recent analysis of Hume’s skepticism in Parusnikova (2016). His more classical sources for the skeptical reading are T.H. Green (editor of the most widely used four-volume edition of Hume’s works until the mid-20th century; (Hume, Green, & Grose, 1874)) and F.H. Bradley for 19th century sources, both idealist philosophers and proponents of a view that sees Hume as a radical sceptic, a “*reductio ad absurdum* of empiricism”, following in the footsteps of Reid and Beattie in the 18th century. A radically skeptical view, which we remember, also characterizes, for example, the reading of Hume by John Stuart Mill or Bertrand Russell in the transition from the 19th to the 20th century.

⁷ I cannot set out or evaluate here the whole of Hume’s epistemology or his very complex moral and political philosophy, but I will only say that Hume’s skeptical version goes directly against the philosophical project already expressed in the subtitle of the *Treatise* (1739-1740): “an attempt to introduce the experimental method of reasoning into moral subjects”. The

Rawls's approach to Hume's thought, however, is based on this classical skeptical-naturalistic reading, which will shape his interpretation of Hume's theory of rationality and will generally be the origin of the attribution of a pure and elementary instrumental theory of practical reason to Hume.

But despite these problems, Rawls' argument deserves close scrutiny. Rawls posits that Hume's scepticism arises from the observation that reason, when untempered by custom and imagination—essentially the benign principles of human nature—tends to self-destruct. He asserts that while pure scepticism is untenable for living a practical life, individuals naturally revert to their beliefs shaped by custom and imagination after philosophical contemplation. Hume's scepticism, particularly in moral philosophy, is thus depicted as a defining aspect of a way of life that contrasts starkly with traditional religious beliefs. This alternative lifestyle, which Hume appears to embrace fully, is viewed as beneficial for moral character and social coexistence. (Cfr. Rawls, 2000, p. 14)

So Rawls distinguishes between two basic types of scepticism, identifying a radical form associated with theoretical philosophy and a moderate form linked to practical philosophy.⁸ He contends that only the immediate contents of the mind—impressions and ideas—resist theoretical scepticism, and that once one moves beyond the confines of custom and imagination, all ideas become subject to doubt. This is where theoretical reason plays a crucial role in liberating individuals from metaphysical ideas and their practical implications.

In his very critical and interesting analysis, Rawls suggests that Hume's scepticism does not yield a coherent (all-dimension) theory of practical rationality. Instead, he argues that Hume's philosophical position emerges from an unexamined equilibrium between his philosophical reflections and his psychological dispositions. This suggests a fundamental irrationalism within Hume's thought, where his conclusions about ethics and justice lack a normative foundation.

Rawls goes on to make the central argument that Hume's ethical scepticism rules out the existence of a substantive theory of practical rationality. He defends that any theoretical framework that Hume might propose would necessarily be descriptive rather than normative and would merely reflect how people typically act rationally (these means are usually related to these ends, but it could be otherwise). In relation to politics, Rawls argues that Hume's scepticism undermines the normative scope of his theory of justice. This theory is actually based on a utilitarian rule schema of an evolutionary and diachronic nature (the 'moral' institutions need time and a clear idea of the utility of their functioning), which could be seen as a critique of contractarianism, i.e. of the idea that the emergence and development of

social institutions and politics are based on a synchronous pact or contract. In spite of the complexity of this approach, Hume's view is portrayed by Rawls and many others as inherently conservative, reflecting the status quo of political institutions at a particular point in time, which is a consequence of his overarching (and unreal) skeptical-naturalist standpoint.

By those reasons, Rawls invites us to consider whether Hume possesses a genuine theory of practical reason. The complex interplay between scepticism and naturalism in Hume's thought, as interpreted by Rawls, raises significant questions about the foundations of ethics and political philosophy and the scope of the Humean theory of rationality, that tends to be read as very short and limited.

But the more we delve into the details of David Hume's supposedly imperfect theory of practical reason, the more details of its subtlety and, more importantly, its increasing normative dimension emerge, to the misfortune of his interpreters and the delight of his readers even if his indirect conception of normativity does not always coincide with a more conventional interpretation of rationality in ethics, the social sciences and economics, i.e. a view of rationality very focussed on drawing direct and immediate consequences of a normative nature (generally reformist) from the totality of facts and data about the nature of man in society.

This attribution of normative ineffectiveness to Hume, understanding normativity as a direct link between a reason and an action, was already to be found in a critic who was almost a contemporary of Hume and was very annoyed by the difficulty (characteristic of Hume's evolutionary and gradualist theory) of obtaining in an immediate and simple way a programme of reform based on his analysis of the criterion of utility as the support of justice. As can be seen in Jeremy Bentham's polemic against Hume, the theory of reason sometimes wants to move too quickly into practise, although this is not a problem for which we can hold Hume directly responsible.⁹

3. Outline and discussion of a hypothetical Humean theory of practical rationality

a. Regarding the existence or absence of Hume's theory of practical rationality and the lack of normative purposes.

Rawls' basic thesis is that Hume does not formulate a theory of practical rationality in the strict sense. Such a theory should be normative and reflect on the goals of action in order to justify them. Instead, Rawls claims that Hume only thinks descriptively about the practical adjustment necessary to achieve certain ends, but not about the ends themselves. Rawls refers to this seemingly purely empirical analysis when

empiricist and probabilistic reading of his philosophy seems much more coherent than his radically skeptical reading, although it must be acknowledged that a certain will on Hume's part to shock and provoke may lie behind the generally skeptical reading of his entire philosophy, which was common in his time and which continued into the 20th century and is still in use today.

⁸ I believe that Hume distinguishes between radical skepticism (Pyrrhonism), which he rejects, and moderate skepticism, which he endorses, but this distinction does not, in my view, correspond to the distinction between theory and practice. In his view, moderate skepticism should apply to both dominions, the one and the other. And radical skepticism must be rejected in both areas.

⁹ Cfr. On the polemic between Bentham and Hume (Tasset, 2018) and (Tasset, 2022).

he claims that Hume has no normative theory of reason but only an instrumental one. (Rawls, 2000, p. 37)

From Rawls' perspective, this means that Reason cannot influence the passions, that is, the realm of motivation, but only our judgments, specifically in the realm of understanding moral situations, moral adjustment, and nothing more.

To support this claim, Rawls refers to the "official conception of rational deliberation" in Hume, which he believes is articulated in the *Treatise* (T 2.3.3.), overlooking the evident evolution between this work from 1739-40 and the *Enquiry concerning the Principles of Morals* (EPM) from 1751.

Hume provocatively stated in the first of these works: "Reason is, and ought only to be the slave of the passions, and can never pretend to any other of office than to serve and obey them." (T.2.3.3.4. SB 415)

In the second work, however, he also points out precisely that there is a fundamental interaction between reason and feelings; both are fundamental for evaluative judgements:

"...reason and sentiment concur in almost all moral determinations and conclusions. The final sentence, it is probable (...) depends on some internal sense or feeling, which nature has made universal in the whole species. (...) But in order to pave the way for such a sentiment, and give a proper discernment of its object, it is often necessary, we find, that much reasoning should precede, that nice distinctions be made, just conclusions drawn, distant comparisons formed, complicated relations examined, and general facts fixed and ascertained." (EPM, 1.9. SBN 173)

In any case, even the position of the *Treatise* on which Rawls focuses, when read in a wider context, is much more complex than it seems, as Hume affirms that there can be, and indeed is, an interaction between reasoning, deliberation, and passions.¹⁰

As Rawls points out by paraphrasing Hume, deliberation, the process of practical reasoning, can correct our passions in at least two ways:

1. It can eliminate one desire and replace it with another, thereby changing our passionate configuration, providing data about the connection between means and ends (instrumental use of reason).
2. It can provide relevant information that we did not have or did not see regarding the nature of things that function as causes or objectives for the passions, that is, on ends and goals (evaluative use of reason).

After which, Rawls interestingly notes that in a certain sense, which contradicts his idea that Hume lacks a theory of practical deliberation, "the process of deliberation may affect the system of the passions in other ways that are not merely correcting

mistaken beliefs but are more constructive." (Rawls, 2000, p. 33)

The forms described by Rawls brilliantly are: specifying, scheduling, and weighting. (Rawls, 2000a, p. 53)

The process of "specification" determines an indeterminate desire.

The process of "scheduling" or "sequencing" arranges the satisfaction of several conflicting passions in temporal order.

And the process of "weighting" establishes a certain order of importance among them:

"Perhaps we face a conflict between final ends and there is no way to schedule them, or to render them more determinate, so as to avoid the conflict. In this case we must assign weights, or priorities, to our ends. (...) Call this *weighting* final ends." (Rawls, 2000, p. 33)

This last dimension seems particularly important for understanding the possibility that Hume pursues a normative and not just a descriptive-instrumental approach. The ends, e.g. social and political, but also individual, sometimes need to be weighed, compared and balanced, and in some cases even prioritised, and in this task the role of reason is crucial. Sometimes not all great goods are possible at the same time, so that it is an essential normative task to resolve, for example, the possible conflict between freedom and security, between democracy and freedom, or finally between equality and justice. This task of establishing a hierarchy of the great principles, purposes or ends goes beyond the mere empirical examination of common practice and is so important that Rawls himself includes it as a fundamental element in his own theory of justice. So there seems to be no good reason to regard this issue as truly normative in Rawls's theory and not to consider it in the same way in Hume's argument.

All this requires a rational endeavour that is clearly substantive and normative and that cannot be approached from a purely instrumental conception of reason, which would merely provide us with logical and empirical information about the relationship between the great ends but would not be able to achieve a determination of the lexicographic order (to use Rawls' own terminology) of these same great principles. But Hume urges us to do so, he does so, surpassing the purely empirical-instrumental approach, as for example when, on the question of the right to rebellion (hypothetical conflict liberty-self-interest-security-stability), he gives absolute priority to the general and individual interest in the face of the possible instability of current power.¹¹

In any case, for Rawls, all this interesting and complex work of rational action, although not necessarily dominant, of reason over the passions, does not form part of that Humean true theory of practical rationality whose existence he is tracing and evaluating. "Observe that he is not presenting an account

¹⁰ There is a fair amount of recent work on "Humean rationality" that generally accepts the ascription of a simple instrumentalist reading of Hume's theory, not questioning this reading in general, but defending the potential of the emotionalist version of behaviour (a kind of internalism, reasons must be connected to inner emotions to become motives, to have motivational force) to renew the discussion of the model of rationality in some way, but without questioning this basic instrumentalism. See for example (Wallace, 2023, p. 16); (Setiya, 2016); and (Smith, 2004).

¹¹ For example, in the essay "Of Passive Obedience", part 2, Essay 13, (Hume, 2021, pp. 347-349).

of rational deliberation understood as normative. Rather, he is saying how, psychologically, we do deliberate.” (Rawls, 2000, p. 38)

This does not prevent him from titling an entire section of his exposition on this topic with the appealing headline for someone interested in practical deliberation and the theory of rationality: “Deliberation as Transforming the System of Passions” (Rawls, 2000, p. 40 and ff.), which is indeed how Hume conceives it, in interaction with the passions, and as one of the mechanisms for their transformation.

Hume thus accepts that deliberation can have an effect on the passions: serene or calm passions, endorsed by reflection and supported by resolution, can control more violent passions.

However, one should not make the mistake of deducing from this that passions are fixed in intensity and influence, and that deliberation directs them. The Humean model understands, rather, through resolution or “strength of mind”: “a virtue built up by custom and habit.” (Rawls, 2000, p. 40). Thus, deliberation is an activity, not a state, and in Hume, who uses a language of virtue theory based, however, on utilitarianism (Crisp, 2005; Spanish trans. by José L. Tasset in Crisp, 2020), the activity of deliberation is linked to the virtues that pertain to the control of violent passions, primarily with what is referred to as “strength of mind.”¹²

However, Hume considers that deliberation, by itself, and this is the key, cannot eliminate or awaken passions.¹³ What reason can do is interact with the passions in the following way, always indirectly: “once serene passions have taken control of the powers of deliberation, they can shape our character to ensure strength of mind.” (Rawls, 2000a, p. 61)¹⁴

That is to say:

(...) “Hume allows that, in certain circumstances, strict reason [logical consistency and empirical adjustment to anticipated probability, JLT] and the calm passions—which we mistake for the operations of reason— together with custom, habit, and imagination, can shape our character over time, sometimes rather quickly.” (Rawls, 2000, pp. 42-43)

In any case, this process of remodelling, of practical and motivational change, which is consistent with granting an important role to deliberation or

practical reasoning, does not awaken for purely intellectual reasons, but for passionate ones. This is characteristic of Hume, even if it does not align with what Rawls, and other critics, expect from a theory of deliberation.

Now, can this strength of mind and the process of adjustment and balance between reason and passions lead to an authentic theory of practical rationality? Rawls seems to think not at all times, but his argumentation, brilliant, very interesting, and detailed, seems to paradoxically demonstrate the contrary.

First, Rawls argues in a very complex way, and somewhat obscure at times, about the internal structure of a possible Humean theory of this kind, emphasizing the differences between Hume and his own Kantian-rooted theory; afterwards, he will do the same with the judicious spectator, the most difficult point to maintain, since not only does this theory seem to correspond to an authentic normative theory of practical rationality, but Rawls himself admits that it represents a truly original intuition of Hume’s as an explanation of the central features of the moral standpoint.

b. The judicious spectator: a plea for a true practical reason theory?

According to Rawls, the idea of the “judicious spectator” “is ‘one of the most important and interesting ideas in moral philosophy.’” (Rawls, 2007, pp. 239-240) And he adds: “It appears in Hume for the first time” (Rawls, 2009, p. 240).¹⁵

This idea emerges in the context of Hume’s discussion of ‘sympathy’ (as it is called in the *Treatise*) or the so-called ‘principle of humanity’ (in the *Second Enquiry*).

The key to our ability to formulate moral judgments about others, or even about our past or future behaviour, the real key to moral and practical rationality, lies in sympathy, understood as the innate capacity of the human species under normal circumstances to feel what these “other subjects” feel (including ourselves in a time or space other than the present).

Now, even for reasons that are certainly innate to the species, this ability to sympathize, to empathize with the feelings of others and thus to identify with their interests, is limited both spatially (the greater the distance, the less the sympathy) and temporally (the greater the temporal distance, the less the sympathy).¹⁶

¹² Rawls provides a very interesting *ontogenetic explanation* of this central Humean concept of “strength of mind” when he says that: “Deliberation, like any other activity, is something we must *learn* to do. It involves forming certain conceptions, going through various steps; it also involves the *imaginative rehearsal* of the consequences of adopting various alternatives, and so on. As we gain *practice*, we do it more easily, and the benefits of deliberating, as judged by our success in fulfilling our calm passions and more basic interests, are greater. Thus facility in deliberation is *rewarded* and this, in turn, gives rise to a stronger inclination, a stronger tendency to engage in deliberation because we enjoy the moderate (not too hard or too easy) exercise of our facility. (It may be odd to speak of *enjoying* deliberation, so let’s say instead that we overcome an *aversion* to it, to the mental exercises it requires.) Further, we learn *when* deliberation is called for, and we come to appreciate its advantages.” (Rawls, 2000, p. 41)

¹³ Cf. the brilliant analysis of this topic contained in (Sen, 2011); Sen, against Rawls, considers that Hume has an interesting theory of practical rationality.

¹⁴ The paradox of this process is that strength of mind seems to require that the individual is in some sense predisposed to that virtue, just as rationality requires a minimum predisposition to it or, what is the same thing, more generally, every virtue requires a natural and to some extent involuntary basis. This structure of argument is characteristic of Humean naturalism and is far from justifying a skeptical position: External reasons for action would ultimately, and in a few very basic cases, rest on internal reasons for action, that is, on motives for action that can interact with reason but are not directly controlled or generated by it.

¹⁵ In the EPM (9.1.6) that expression does not appear but rather a much clearer one, the “common point of view.” This point of view is structured around the “principle of humanity”, in which “every man, in some degree, concurs”, and which is the concept that in general in this work replaces that of sympathy (dominant in the *Treatise*).

¹⁶ A very interesting recent analysis of the role of natural empathy, its limited and restricted nature and the possible consequences of its extension by artificial means can be found in Lara (2017).

Sympathy thus enables moral judgement and at the same time hinders its objectivity; as long as we do not extend it, we cannot be impartial. This is where the “judicious spectator” comes into play, because we humans are also able to compare our own point of view, the particular position that each of us occupies in time and space, our particular interests, with some more fixed and generalized points of view:

“In order...to...arrive at a more *stable* judgment of things, we fix on some *steady* and *general* points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation.” (T 3.3.1.15; SBN 581-2)

Given this original argument of Humean moral philosophy, Rawls again raises the question of whether Hume is somehow elaborating and proposing a theory of practical rationality here, or whether he is confining himself to describing what people actually do, straightforwardly acting, as Hume often says, as an anatomist (describing empirical details) rather than a painter (abstract understanding of reality). In any case, the idea of the judicious spectator, of the common point of view, seems so original and correct that Rawls’s position in this case is negative, but with hesitation: in his view, Hume has no such theory of practical reason, but he *might have*:

“Let’s now consider whether Hume’s account of the judicious spectator and its epistemological role includes a conception of practical reason, or whether it is instead an account of the psychological processes whereby our moral judgments are expressed. In Hume’s mind, I believe, it is a psychological account. There are, however, passages that suggest how this account might become, if pressed, a conception of practical reasoning”. (Rawls, 2000, pp. 96-97)

Now, let us briefly follow Hume’s argument to see whether it really does, or at least could, give rise to a theory of rational deliberation, an authentic normative theory of practical reason, and not just a mere argument about the need to properly adjust our means and “given” ends right.

What Hume is saying is that when we formulate a moral judgement, i.e. a judgement about good or bad, right or wrong, just or unjust, we are making a judgement about the qualities of character (for Hume the proper object of evaluation) or of actions (as effects or signs of these character qualities, hence the indirect object of evaluation), “we are considering them from a suitably general or “common point of view,” the point of view of the “judicious spectator,” without any reference to our own interests” (Rawls, 2007, p. 185).

What strikes Hume is the fact that when we humans evaluate the same character traits or actions in terms of our own specific interests, we do not reach agreement, but we do when we appeal to what he calls the “principle of humanity” in the *Enquiry concerning the Principles of Morals* (9.1.6. the clearest example), which means that “our judgments are guided by the tendency of these actions or qualities or institutions to affect the general interests of so-

ciety or the general happiness of society.” (Rawls, 2007, p. 185)

So, in accordance with the Humean approaches to the interaction between passions and reason, we thus find that what makes impartial moral judgement possible, and is therefore general and at least relatively universal, is a passionate disposition with a basic cognitive function crucial for moral and practical reasoning, even a precondition for that: the feeling of humanity, “our being able to take up and to imagine ourselves into the point of view of the judicious spectator.” (Rawls, 2009, p. 239)

Explained in this way, it seems clear that Hume’s position would lead from the fact of our sense of humanity to the use of this sense as a disposition of thought or judgement (a principle of human nature), as a mechanism for adopting an impartial position that ultimately allows us to always keep in mind the utility of behaviours or institutions for the welfare of the same humanity that we perceive as common at the origin. This seems to be an argument that corresponds to a more than modest theory of practical reasoning, and to the extent that it is fundamental to the subsequent construction of institutions of common life, it is also clear that it could encompass normative and not merely descriptive functions.

Given these arguments, Rawls will once again claim (cfr. Rawls, 2007, pp. 186-187) that Hume’s naturalistic scepticism can only lead to a purely psychological and descriptive approach (this is what people do) that lacks normative scope (there is no option for what they ought to do) and, worse, if it had a remote normative claim, it would be of a conservative kind, of mere verification and permanence of the pre-existing ethical, political and normative situation.

Ultimately, the problem seems to be that the simple, sceptical and naturalistic interpretation of Hume’s general philosophy and his moral and political philosophy in particular seems to characterise the whole reading that Rawls and countless interpreters of Hume from the contemporary social sciences have defended, as we have shown.

But Hume’s thought resists this simple interpretation because it moves precisely between descriptive and normative domains and considerations. The only source of rational normativity is human passions, the real interests of individuals and society. From this we cannot *deduce* a normative discourse, as all rationalist philosophies have intended. Rather, logical evidence and statistical probabilities form the substrate, the only possible one, for the normative discourse, which will be provisional, indirect, verifiable and purely probabilistic.

This fallibilist scheme may be unsatisfactory to more rationalist thinkers associated with the Kantian tradition, such as John Rawls, but it could, as we shall see, have enormous potential, for example, to construct a proposal, formally more a post-Humean development than an explicitly Humean one, concerning the potential conflict between morality and rationality arising from what Hume calls “knavery” or “rational breach of rules”, a practical situation first described in empirical terms but with a normative solution rooted in the crucial importance for Hume of institutional design. But how does this example of an indirect way between empirical, instrumental and foundational normative reason feel?

4. A hypothetical normative development of a post-Humean nature.

We have seen how, paradoxically, the extensive and detailed analysis that John Rawls undertakes in his lectures on the history of moral philosophy regarding the ethical and political thought of David Hume, in order to demonstrate that Hume does not possess a theory of normative rationality but merely a descriptive one (another way to label it as purely instrumental), that is, focused on achieving the most perfect possible adjustment between means and ends, nevertheless demonstrates that, especially in the realm of the original idea of the impartial observer or “judicious spectator”, Hume seems to develop a theory with evident normative implications, even though he always attempts to address the normativity of reason in a characteristically indirect, we might say, reluctantly moralizing way: his analysis of descriptive, naturalistic components must, by themselves, open the door to normative conclusions in a very open manner, and never in a deductive, closed, or exclusive way.

A brilliant example of this can be found in his analysis of the problem (not a paradox, in his opinion) of the free-rider, as he famously explains at the end of his “Inquiry into the Principles of Morals” (IPM) which he illustrates with the example of the “sensible knave.” Strictly speaking, the “sensible knave” is a rational subject who has a good reason not to comply with all the rules of justice and to gain selfish benefits because everyone else already complies generously with those same rules. We cannot use a direct rational argument against him because what he does is rational. However, we can use other indirect rational arguments, such as the preventive institutional design of these mismatches with the threat, for example, of imposing punishments of such magnitude that they dis-incentivize any kind of selfish option if detection occurs.

In any case, Hume’s analysis of this issue once again leaves the door open to a much more complex normative development based on the simple manifestation, seemingly purely descriptive, of a conflict between morality and rationality. In reality, Hume’s theory of practical rationality suggests a much more original and novel idea: it is not so much about distinguishing between instrumental or normative rationality, selfish or cooperative, but between different levels of morality and rationality.

To conclude, we will briefly attempt to sketch what such a theory might consist of, keeping in mind that it is more of a post-Humean development than a position that has literally been developed by Hume himself.

a. Morality and Practical Self-Restraint.

This development of an argument for compatibility between morality and rationality, constructed from Hume’s theory of institutions, requires a somewhat

restricted use of the term “moral,” which we will understand basically as “a set of self-restrictions,” that is, self-imposed limitations on the complete universe of actions available to a moral subject. Morality will consist of autonomously and voluntarily restricting that universe. Although a subject may want to do everything, they decide to do only certain things and not others. This set of self-restrictions does not have to be fixed, as the majority of deontological or “principled” ethics seem to defend, but can depend significantly on the context, primarily defined by the behaviour of the other individuals with whom we interact.

This set of self-restrictions, initially (Morality type 1 or M1), is based, according to Hume’s theory, on an evaluation of our own interest that is directly linked to the interests of others but close to us (family, friends, neighbours, etc.). Therefore, it rests on a non-selfish rationality (there is no primacy of exchange, and we could even speak of *Tit for Tat* at a total loss), which imposes itself over selfish rationality (exchange dependent on immediate, clear, and distinct reward; ordinary *Tit for Tat*). From this situation, when we move away from the immediate social sphere, we transition to an evaluation of our own interest in which it is linked to strangers (Morality type 2 or M2), and in which, however, two situations or scenarios can occur:

(M2a) Either there is an institutional design that makes it rational to continue using the same set of self-restrictions from (M1), and therefore we can continue to prioritize medium-to-long-term rationality over short-term rationality, that is, we remain altruistic instead of selfish, and this does not make us irrational because, collectively, that same cooperative or altruistic position is adopted by all subjects and is incentivized, reinforced, and even rewarded formally or informally by the institutions (and dis-incentivized and punished when it does not occur) (although with a much broader dimension facilitated by our ability to feel sympathy for individuals who are not close or immediate), or:

(M2b) There is a failed institutional design, and the set of self-restrictions (M1) no longer holds, so we must prioritize short-term, or selfish, rationality to remain rational, at least until the design changes.¹⁷

b. Morality by Levels: A Post-Humean Thesis on the Conflict between Morality and Rationality.

Hume’s thesis on moral evolution from a natural realm (driven by natural passions and approved through natural virtues) to an artificial realm (driven by indirect and derived passions, approved by artificial virtues) has an important normative consequence, that is, from its empirical analysis we can get as a result a probable argument that biases, in a non-dogmatic way, on a solution or set of solutions.

M2, or conventional or artificial morality, centered on Justice, will be based not only centrally but exclusively on Utility.

¹⁷ Alfonso Palacio-Vera very aptly suggests that (M2b) and (M2a) could correspond to two possible situations in a game like the “prisoner’s dilemma”. (M2a) would correspond to the socially optimal equilibrium, while (M2b) would correspond to the socially suboptimal equilibrium. In any case, this dilemma seems to serve to analyze whether or not there is cooperation between individuals, whereas the “Humean matrix” representing cooperation refers more to the development of the interaction (and iteration) between individuals and institutions or regulatory contexts in which the trust that serves as the subsequent basis for cooperation is established.

Not so M1, or natural morality, which is based on naturalistic considerations (not only in the sense of natural but also immediate, direct, requiring little or no calculation), which may derive its approval from Utility or not. In this realm, it makes sense to assert, as Hume does to Bentham's scandal, that Utility is "one foundation of morality" but not "the sole foundation of morality."¹⁸

And indeed, there will be moral actions in the realm of natural socialization (familial or multi-familial societies) or S1, related to character, demanding approval and generating virtuous patterns of behaviour, therefore worthy of praise and imitation, that are not based on direct Utility for the individual: for example, the non-reciprocal altruism of a parent towards their children or their elders.

From this perspective, Hume's thesis regarding the articulation of naturalism and utilitarianism is not only brilliant but allows for the explanation and even resolution of a classic ethical problem: morality and rationality do not have to oppose each other if they evolve dynamically.

Thus, my basic thesis will be, following Hume and at the same time going much further than what he stated, that the type of morality must be adapted at each moment to the form of social interaction, based on whether or not it is founded on reciprocity and the generalization of individual virtuous behaviours, giving rise to the following possible combinatory (with M as the type of morality and S as the type of society):

M1	M1	M2	M2
S2	S1	S1	S2
(-)	(+)	(-)	(+)

In the first situation, we try to operate with a natural, direct morality in a complex social environment and fail; in the second, we operate with natural motives and morality in a nearby and direct social environment and succeed; in the third, we function with an artificial, complex morality based on utility in a nearby and direct social environment and fail (our family members flee at the attempt to apply a *tit-for-tat* morality to them); finally, in the last stage, we apply a complex morality of utilitarian roots (of the rule, in my view) in an environment of distant and complex sociability and succeed, because we mainly avoid the paradox of the 'sensible knave', the clever thief who is always looking for a moral idiot—or many—to bargain with and deceive if he manages to get them—or many—to continue practicing a direct, solidary, and altruistic morality (it can be reciprocal, but this reciprocity is not a *sine qua non* requirement) that excludes *tit for tat* in an environment where the key is that everyone practices a morality of a radical reciprocity not only binary but generalized.

c. Moral Multidimensionality.

As can be observed in our brief post-Humean development, the theory on the evolution of institutions by David Hume, initially and in a first glance purely empirical, psychological, and instrumental, has important normative consequences.

First, and primarily based on the distinction between natural and artificial virtues, it opens the door to a complex understanding of morality by levels.

Second, it allows us to understand that these levels are differentiated by the following factors:

- Different types of passions operate within them.
- In accordance with this, two different types of rationality also operate, and
- These different types of rationality correlate with different types of utility.

And third, it provides the basic intuition to develop the idea that when an adequate adjustment occurs between these factors, there will be no contradiction between morality and rationality; when, on the contrary, a mismatch occurs, a dissonance between morality and rationality arises, and a quick intervention on the design of social, political and economic institutions is required, not in an indeterminate sense, but in the line of restoration of the balance between morality and rationality, or between the proper levels and senses of both instances of human reason.

Just as complex society arises from very elementary familial or endogamous structures, morality based on justice also arises from a quasi-natural morality based on natural passions such as love, although these passions may no longer suffice in an advanced and complex stage of moral and institutional evolution.

In any case, Hume's basic thesis on the articulation between natural and artificial passions lays the groundwork for a very apt solution to the potential conflict between morality and rationality, or more precisely, between different types of morality, defined by the use of different types of rationality. Artificial passions are founded on natural passions. In the realm of very basic morality, natural passions suffice. In the realm of more complex morality, artificial passions are needed, which, however, are rooted in the natural motive of interest, even if they may serve it in a much more remote and distant manner. Our morality and our rationality can and should change as our social interaction becomes more complex and our natural motives for action become more remote. Nevertheless, these motives are the original source of morality, and if they are constantly and systematically contradicted, as the problem of the 'sensible knave' suggests, we will need a normative intervention in the way we have defended.

So, to conclude, we believe we have shown that Hume never held a purely instrumental theory of reason, just as his scepticism regarding its practical functions is not radical, which allows for a better understanding of his normative approach to moral, political, and practical rationality, and even enables us to draw significant normative conclusions from his work, particularly and very relevantly from his analysis of the problem of the 'sensible knave' and the hypothetical conflict between morality and rationality that can be overcome with the development of a theory of levels of practical rationality.

¹⁸ For an analysis of Hume and Bentham differences on the use of Utility cfr. (Tasset, 2019).

5. Conclusions

As we explained at the beginning of this paper, it has become customary to regard instrumental rationality as a fundamental element of the so-called standard conception of practical rationality in moral and political philosophy and in the social sciences, especially in economics. The rational subject would be the one who chooses an optimal fit between the means at his disposal and the ends or preferences he pursues. The second element in this description of instrumental rationality would be a purely historiographical judgement, according to which this conception has its origins in Hume's arguments in the *Treatise* of 1739-40 about the limited character of reason in morality and the primacy of the passions. This standard attribution has become so common that there are very few places where we find a detailed analysis of Hume's thesis and whether he really holds what is attributed to him. One exception is the detailed examination of David Hume's moral and political philosophy undertaken by John Rawls in his extraordinary *Lectures on the History of Moral Philosophy* and *Lectures on the History of Political Philosophy*.

Given the respect and interest Rawls has for Hume's theory of practical reason, and given the fact that, despite adherence to the standard version (Hume has nothing more than a descriptive, psychological, purely instrumental and non-normative theory) of Hume's *Ethics and Politics*, we have started from his account in order, in short, to contradict his interpretation by interpreting his own arguments and to try to show that Hume has a theory of practical reason with important normative implications although he also represents a vision of indirect moral normativity

that is very far from the deductive and direct canons defended by other philosophies, especially that of Rawls himself, which is based on his Kantian inspiration. We therefore believe that we have shown that David Hume's theory of moral and practical rationality has great potential for current discussions on human action, even if this requires reading it away from the stereotype that anchors it in this purely instrumental and mechanical conception of rationality and its apparently passive submission to the emotional elements summarised in the concept of "passion".

Rawls shows great respect for Hume's theory of practical reason, even if he ultimately adheres to the standard reading of this theory: Hume has nothing more than a descriptive, psychological, purely instrumental and non-normative theory. We generally disagree with this interpretation and have tried to show that Hume has a theory of practical reason with important normative implications. But Hume also advocates an indirect vision of moral normativity that is very far from the deductive and more direct canons defended by other philosophies, notably Rawls himself, who relies on Kant's views. We therefore believe that we have shown that David Hume's theory of moral and practical rationality has great potential for current discussions on human action, even if this requires reading it away from the stereotype that anchors it in this purely instrumental and mechanical conception of rationality and its apparently passive submission to the emotional elements summarised in the notion of "passion".

Hume seems to be saying more than this and we need to pay attention to his global theory of practical rationality and listen up.

Bibliographic references

- Burnyeat, M. (1983). *The Skeptical tradition*. Berkeley: University of California Press.
- Crisp, R. (2005). "Hume on Virtue, Utility, and Morality." In S. M. Gardiner (Ed.), *Virtue ethics, old and new* (pp. 159-178). Ithaca, N.Y.: Cornell University Press. Spanish translation by José L. Tasset in Crisp, R. (2020). "Hume: sobre virtud, utilidad y moralidad." *Télos. Revista Iberoamericana de Estudios Utilitaristas*, 23(1-2), 9-35. <https://doi.org/10.15304/t.23.1-2.6781>
- Fogelin, R. J. (2009). *Hume's skeptical crisis: a textual study*. Oxford, New York: Oxford University Press.
- Hume, D., Green, T. H., & Grose, T. H. (1874). *The Philosophical Works*. [S.I.]: Longmans.
- Hume, D., Norton, D. F., & Norton, M. J. (2007). *A treatise of human nature: a critical edition*. Oxford/ New York: Clarendon Press/ Oxford University Press.
- Hume, D., & Beauchamp, T. L. (1998). *An enquiry concerning the principles of morals. Critical Edition*. Oxford: Clarendon.
- Hume, David (2021). *Essays, Moral, Political, and Literary. A Critical Edition*. Edited by Tom L. Beauchamp & Mark A. Box. Volume 1. Oxford: Clarendon Press.
- Lara, F. (2017). "Oxytocin, Empathy and Human Enhancement." *THEORIA. An International Journal for Theory, History and Foundations of Science*, 32(3), 367-384. <https://dx.doi.org/10.1387/theoria.17890>
- Parusnikova, Z. (2016). *David Hume, sceptic*. New York, NY: Springer Berlin Heidelberg.
- Railton, P. (2007). "Humean Theory of Practical Rationality." In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory*. Oxford University Press.
- Rawls, J. (2000). *Lectures on the history of moral philosophy*. Cambridge, Mass: Harvard University Press.
- Sen, A. (2011). "The Boundaries of Justice (David Hume and our world)." *The New Republic*, 29 (December).
- Setiya, K. (2016). "Hume on Practical Reason", in *Practical Knowledge: Selected Essays*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190462925.003.0013>
- Smith, M. (2004). "Humean Rationality." In A. R. Mele & P. Rawling (Eds.), *The Oxford Handbook of Rationality*. Oxford University Press. <https://doi.org/10.1093/oxfordhdb/9780195145397.003.0005>
- Tasset, J. L. (2011). "De ladrones y reglas (Una visión del problema del "Sensible Knave" desde un utilitarismo de la regla atemperado)." *Daimon, Revista Internacional de Filosofía* 52(enero-abril): 117-140
- Tasset, J. L. (2018). "Instituciones, evolución y delincuencia racional: hacia una perspectiva posthumeana." *Araucaria: Revista Iberoamericana de Filosofía, Política, Humanidades y Relaciones Internacionales* 20(40): 349-369.

- Tasset, J. L. (2019). "Bentham on 'Hume's Virtues'." Happiness and Utility. *Essays Presented to Frederick Rosen*. G. Varouxakis and M. Philp. London: UCL Press: 81–97. <https://doi.org/10.2307/j.ctvf3w1s5.9>
- Tasset Carmona, J. L. (2021). "En torno a la lectura rawlsiana de la filosofía moral de David Hume." *Anales de la Cátedra Francisco Suárez*, 55(0), 155-182. <https://doi:10.30827/acfs.v55i0.15571>
- Tasset, José L. (2022). "La ética utilitarista de Jeremy Bentham: del Ser al Deber ser." *Télos* 24(1-2). <https://doi.org/10.15304/telos.24.1-2.8085>
- Wallace, R. J. (2023). *Normativity and the will: selected papers on moral psychology and practical reason*. Oxford scholarship online. <https://doi.org/10.1093/oso/9780199287482.002.0001>.