ATHENA

CRITICAL INQUIRIES IN LAW, PHILOSOPHY AND GLOBALIZATION

Fictional Mechanical Minds

On the Relationship Between Assumptions of Rationality and Conceptions of Government

PAOLO CAPRIATI

Postdoctoral Research Fellow in Philosophy of Law, University of Palermo (Italy)

✉ **paolo.capriati@unipa.it**

https://orcid.org/0009-0002-6027-441X

ABSTRACT

If a machine were considered more rational than any other agent, would we entrust it with the government of public affairs? This paper aims to examine the relationship between rationality and political justification. The thesis whose soundness I will verify is: (T1) Depending on which subject's (greater) rationality is assumed, different conceptions of government emerge. The link between assumptions of rationality and political legitimacy implies another thesis: (T) Under certain conditions, to legitimize a subject to decide on matters of collective interest, it is necessary to assume that the deciding subject is rational. Firstly, the terms of the issue must be clarified: (1) the subjects who confer legitimacy on a government; (2) the assumptions of rationality; (3) rationality; (4) the decision-making subjects; (5) machines as decision-making subjects; (6) the conceptions of government. Secondly, under what conditions do the assumptions of rationality become essential to legitimize a government's decision on matters of public interest? Finally, starting from three distinct political subjects – (A) the individuals, (B) the individuals conceived as a collective, and (C) the machine – it will be shown how, by assuming their rationality, three distinct conceptions of government take shape that are: (A1) aggregative democracy; (B1) de-liberative democracy; (C1) machine-government.

Keywords: rationality, political legitimacy, decision-making process, aggregative democracy, deliberative democracy, machine-government

ATHENA
Volume 5.2/2025, pp. 46-70
Articles
ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

1. Introduction – Presentation of the Problem

Let us imagine that it were possible to entrust the government of public affairs to a machine. Why should we do so? There are many possible answers to this question. Let us assume, plausibly, that the following argument is given:
We should entrust the government to a machine because a machine can be more rational than any other agent.

Let us therefore focus on the relationship between rationality and political justification.

The thesis whose soundness I will attempt to verify is the following:
(T_1) Depending on which subject's (greater) rationality is assumed, different conceptions of government emerge.

What I will argue is that there is a link between assumptions of rationality and political legitimacy[1] such that, as the subject whose (greater) rationality is assumed varies, the conception of government also changes.

The link between assumptions of rationality and political legitimacy implies another thesis:
(T) Under certain conditions, to legitimize a subject to decide on matters of collective interest, it is necessary to assume that the subject who decides is rational.

The presentation of these theses is divided into two parts.

[1] "Legitimacy" – understood in a descriptive sense – refers to the mere acceptance of a government by its subjects. "Justification", on the other hand, refers to the reasons used to legitimize a given government. For example, government X is legitimate because the governed hold certain beliefs or a particular faith in it (Weber, 1964); government X is justified because there is a specific reason why the governed regard it as legitimate. The distinction between legitimacy and justification is not as clear-cut as it may seem. More precisely, the purely descriptive concept of legitimacy has been challenged, since it does not consider second-order beliefs about legitimacy – that is, beliefs about what is necessary for a given institution to be considered legitimate. According to Beetham, a "power relationship is not legitimate because people believe in its legitimacy, but because it can be justified in terms of their beliefs" (Beetham, 1991, 11). As will be seen, according to my reconstruction, the reason why, in some cases, the governed consider a government legitimate is that they believe government X to be rational. In other words, under certain conditions, a government is legitimate – and justified – if it is regarded as rational.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

In the first part, I will define the key terms of the discussion. First, I will clarify what I mean by “assumptions of rationality” and explain in what sense the law relies on such assumptions for its functioning. Before that, it is necessary to define the concept of rationality itself and specify how it is understood in this context. Secondly, I will clarify what I mean by “conceptions of government”. This notion is closely connected to the model of government and to the political decision-making procedure. I will refer to the political decision-making procedure as the process that leads to decisions on matters of collective interest. It will therefore be necessary to clarify what constitutes such a process, what forms it may take, and how it can be justified. Finally, I will address the possibility that the decision-maker may be a machine, specifying how I use the concept of “machine” and in what sense a machine defined in this way can be considered rational or not.

After defining these essential terms, the second part of the work will investigate the connection between assumptions of rationality and political legitimacy. The aim is to determine the conditions that make it necessary to assume that the political decision-maker is rational. To illustrate the link between assumptions of rationality and political legitimacy, I will present an essential taxonomy in which, as the subject whose rationality is assumed changes, different conceptions of government take shape.

The subjects whose rationality may be assumed, which I will consider, are three:

(A) individual citizens;

(B) individuals conceived as a collective (provided that they interact in a certain way);

(C) machines.

The conceptions of government that arise from these three different subjects are, respectively:

(A_1) aggregative democracy;

(B_1) deliberative democracy;

(C_1) machine-government.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

In summary, the two theses I will try to demonstrate are:

(T) In some cases, assumptions of rationality are necessary to legitimate the subject who decides on political matters;

(T_1) In those conceptions of government relying on the criterion of rationality, depending on which subject's rationality is assumed, different conceptions of government emerge.

2. What are the Assumptions of Rationality, and what Purpose do they Serve?

By "assumptions of rationality", I mean the belief that a certain subject is considered rational. This assumption has been defined as "rationality perfectionism", and it has been suggested that the law supports this assumption by postulating fictitious cognitive abilities of individuals (Ubertone, 2023). The fictitious nature of these abilities is the subject of recent psychological studies that have shown how the mind works in a way that is far from rational – or at least not always rational (Kahneman, 1994; Stich, 1990).

Rationality perfectionism assumes that the rational subject is the individual. The use of this myth is to some extent necessary for the proper functioning of the law. If we did not assume, for example, that individuals can correctly understand the laws, criminal law would lose its *raison d'être, since anyone could claim a lack of understanding of the law as a cause for excluding* guilt.

The basis of this myth has been challenged, as it relies on the assumption that subjects can act rationally – an assumption that is not universally accepted. The psychological literature of recent years seems to go in the opposite direction. Some works in the social sciences have embraced these studies and have developed governance strategies based on assumptions of mind functioning that are far from rational: one example is nudge. Nudge assumes that individuals are essentially irrational beings (Thaler and

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

Sunstein, 2008). More precisely, according to the proponents of nudge, an effective way to influence individuals consists in adopting automatic, unconscious, and emotional mechanisms of the human mind – defined by Kahneman as "System 1" thinking – and not relying exclusively on rational and conscious deliberation, which, requiring a certain cognitive effort, is activated less often – defined as "System 2" thinking (Kahneman, 2011).

Ultimately, nudge suggests that the law should adapt to human nature, taking into account cognitive biases and irrational tendencies, to guide choices more effectively, without limiting itself to assuming an ideal rationality that does not exist in reality (Thaler and Sunstein, 2011).

One issue remains unresolved: which subjects' assumptions of rationality are relevant? The answer seems obvious: they are the same subjects who hold the capacity to legitimize a government. Legitimacy, in this context, is understood as the belief in legitimacy (Weber, 1978, 213) and denotes the adherence or acceptance of the governed toward a given government. In this sense, the subjects who confer legitimacy – as well as those whose assumptions of rationality are relevant – are the governed.

2.1 Models of Government and Conceptions of Government

We have seen how the law's assumptions of rationality work: it is the law that assumes that members of society are perfectly rational. In the case of political legitimacy, we witness the reverse process: it is the members of society who assume the rationality of those who decide.

It is now appropriate to clarify what is meant by "government". I assume that the government is the subject, or the set of procedures, responsible for making decisions of collective interest. Such decisions are made through a political decision-making process. "Political", in this context, can be defined as that which concerns the interests of a given community.

It is therefore necessary to identify the boundaries of the political decision-making process. In ordinary language, we reserve the terms "government" and "political decision" for activities that regulate large groups of people in potentially all aspects of their lives. However, to understand the basic structure of a political decision-making process, it may be useful to consider a toy example involving a small group of individuals rather than a political community, and a decision of very limited scope rather than a fully political one.

Suppose that in a library, it must be decided whether to close the window or not, and three people are sitting in the library who disagree about what to do. In this case, the decision-making process is what enables them to decide on the window – essentially, whether to leave it open or to close it. The possible decision-making processes, in this case, are at least four:

(1) A random process, in which the decision is made, for example, by flipping a coin;

(2) A violent process: after a fight among the three individuals, the strongest decides; here, the decision-making process coincides with physical confrontation;

(3) An electoral process: the decision is made by voting (possibly preceded by debate), generally following the majority rule. In this case, two votes in favour of "open" or "closed" are enough to decide;

(4) A process where a qualified subject decides: the decision is delegated to a particular subject, based on a criterion of competence. For instance, the person sitting closest to the window might be chosen to decide, as they know better the effects of the window being open or closed – such as whether insects might enter.

These four are examples of models of political decision-making processes, or models of government. Each could be justified with different reasons. (1) and (2) do not belong to the instruments typically accepted in Western legal-political culture. While (1) could be justified by the idea that chance should decide, (2) could be justified by an appeal to physical strength. (3) and (4), on the other hand, share the idea that rationality might play a role in determining who decides. More precisely, in (3) those present in the library could assume themselves to be sufficiently rational to decide. In (4), they could believe that among them there is someone capable of making a more rational (or informed) choice than the others – in this case, the person closest to the window. However, (3) could also be indifferent to the principle of rationality and justified solely because, through (3), everyone can take part in the decision, and it is good for everyone to do so.

This brings us to the distinction between "models of government" and "conceptions of government". By conception of government, I refer to the normative structure that underlies a given model of government. (3), for example, can be understood both as a model of government and as a conception of government, if emphasis is placed on the reasons that justify that particular model. In this sense, for each model of government, we can have as many conceptions of government as there are reasons (or sets of reasons) that provide a normative foundation for that model. For instance, in the case of (3), we could have:

(A) a conception of government that justifies (3) because it allows everyone to participate in the political process;

(B) a conception of government that justifies (3) because it assumes that the deciding subjects are rational.

2.2 On the Concept of Rationality

An attentive reader will have noticed that I have not yet provided a definition of rationality.[2] The definition that seems most suitable for our purposes is that of instrumental rationality. According to this view, rationality manifests itself as the use of appropriate means to achieve one's ends (Kolodny and Brunero, 2013). Therefore, instrumental rationality presents itself as a quality of individuals. "I assume that someone is rational" means that I expect that person to act using the means suitable to achieve a certain end.

[2] The concept of rationality should not be confused with that of intelligence. Although both can be regarded as qualities of a subject, intelligence is a more demanding concept. To clarify: a thermostat can act in a perfectly rational way – from an instrumental point of view, by regulating the temperature precisely – and yet it seems evident that it cannot be considered an intelligent agent.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

Let us try to apply this definition to the case of the library window. In this case, the subject whose rationality could be assumed – that is, the one who could be legitimate to decide – is the subject who makes the decisions appropriate for achieving a given end. But who determines this end? One might object that it is precisely the task of the decision-maker to determine the end – and that this end would coincide with the content of the decision itself. However, framing the problem in this way is misleading. The end, in this case, concerns identifying the best possible state of the window – where "best" is a variable that usually depends on the preferences of those present in the library[3] – and not the decision itself. The decision, in fact, is only the instrument through which to achieve the end, namely, determining the best possible state of the window.

2.3 Can we Assume the Rationality of a Machine?

Is it possible to legitimize a machine to decide? If, as we have seen, there is a link between assumptions of rationality and political legitimacy, we must ask whether it makes sense to assume the rationality of a machine. Even before that, it is necessary to clarify what we mean by "machine".

By machine, I refer to an autonomous non-living agent. I will examine these three aspects separately: (A) agent, (B) non-living, (C) autonomous.

(A) According to Luc Steels (1995), an agent:

- is a system (a set of elements that have relationships with each other and with the environment);[4]
- performs particular functions for another system;
- is able to preserve itself.

It has been observed that not all systems are agents. More precisely, agents must be distinguished from "plant-controller systems", which:

[3] There is a hidden assumption in this statement, namely that the concept of 'end' is determined by the desires of those who are present in the room. Probably, this is precisely the appeal of democracy: to communicate to the decision-makers what the group's end is, thereby enabling them to make the decision that allows that end to be achieved.
[4] Steels (1995) observes that agents do not necessarily need to be physically situated.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

> are formed of two essential components, the plant or system whose behaviour is to be controlled, and the controller which measures the state of the plant, and thus aspects of its behaviour, and initiates control actions so as to keep it operating within allowable or acceptable limits of behaviour. Of course, plant-controller systems really have a third component, the environment, but this is usually left in the background in any description of the system (Steels, 1997, 7).

On the other hand, agents – and the environments with which agents interact – shape agent-environment interaction systems (Smithers, 1994). In such systems, there is no plant to control and no controller either. It is the agent that is responsible for initiating and sustaining effective interaction with its environment (Smithers, 1997, 8).

(B) It is necessary to examine the concept of living. Living systems:

> are characterized by exergonic metabolism, growth and internal molecular replication, all organized in a closed causal circular process that allows for evolutionary change in the way the circularity is maintained, but not for the loss of the circularity itself. (…) [Cf. Commoner, 1965]. This circular organization constitutes a homeostatic system whose function is to produce and maintain this very same circular organization by determining that the components that specify it be those whose synthesis or maintenance it secures. Furthermore, this circular organization defines a living system as a unit of interactions and is essential for its maintenance as a unit; that which is not in it is external to it or does not exist. The circular organization in which the components that specify it are those whose synthesis or maintenance it secures in a manner such that the product of their functioning is the same functioning organization that produces them, is the living organization (Maturana and Varela, 1992).

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

For an agent to be considered “non-living”, it must lack even just one of these characteristics.

(C) Greater difficulties in framing arise with the concept of autonomy. Various disciplines – such as biology, law, philosophy, and ethics – have addressed the definition of autonomy. What these definitions have in common is the possibility of “self-law making” (Smithers, 1997). The minimal definition to start from is as follows: an agent is autonomous if it is relatively independent from something else (Steels, 1995). In this case, the independence refers to humans: for a machine to be considered autonomous, it must be independent of its programmer to the extent that its actions are not entirely predetermined by a human agent. This autonomy then translates into the ability to determine its own laws. Phenomenologically, there are two facts indicative of a state of autonomy: the relative impossibility of accounting for behaviours *ex post*, or the relative unpredictability of such behaviours.

Finally, I will not consider intelligence as a useful criterion for defining the machine. In defining the intelligence of a non-living autonomous agent, there are at least three tendencies:

(1) considering intelligent the machine that exhibits behaviour similar to that of humans (Turing, 1950);

(2) considering artificial intelligence not comparable to human intelligence, using concepts – such as consciousness (Penrose, 1991);

(3) considering intelligent the machine whose functioning is not reducible to a mere physical process (Steels, 1995).

Following (1) risks considering intelligent a machine that merely imitates human behaviour. In the case of the chess-playing machine, for example, there are programs carrying out deep searches within the search space, and their impressive performance is therefore no longer regarded as an expression of intelligence (Steels, 1995), but rather as the result of a machine’s ability to draw upon a vast memory. As for (2), in the absence of a clear definition of consciousness, claiming that “intelligence is linked to consciousness” is a fundamentally vague statement. (2), using undefined concepts, excludes *a*

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

priori that a machine can be intelligent. However, the limit of (3) is that, by drawing such a sharp distinction between cognition and action, it seems to overlook the fact that every cognitive process always originates from a physical process – just as, in the case of the human brain, thought takes shape from neural activity.[5]

In the absence of an appropriate definition, it must be concluded that intelligence is not a criterion that, in this context, can be usefully adopted to describe a machine. Therefore, this criterion must be abandoned.[6]

It has been seen that for the law to fulfil its role, it is necessary to assume the rationality of individuals. But is it possible to also assume the rationality of a machine? This question implies a philosophical query of considerable depth: can machines be rational agents?

This issue is related to the problem of the thinking machine. This problem was first presented by Turing in the famous article “Computing Machinery and Intelligence” in 1950. Since “machine” and “thinking” cannot be defined unambiguously, Turing suggests reformulating the question “can machines think” by replacing it with a less ambiguous question: are there discrete state machines that would perform well in deceiving a human observer, making them believe they are conversing with a human being rather than a machine?

Turing’s reformulation shows that the problem should not be confronted head-on but rather approached indirectly. A lateral approach, like the one proposed in “Computing Machinery and Intelligence”, allows avoiding concepts that cannot be easily operationalised. The merit of this reformulation is also that it makes a response possible. In other words, by providing an empirical standard, this question offers a verifiable criterion based on the imitation of linguistic abilities. In this way, the focus shifts to behaviour (which is observable) rather than a supposed internal state of the machine.

[5] Unless one intends to adopt a definition of intelligence that is not physically grounded, assuming a form of intelligence that exists independently of matter.
[6] I have dealt with the topic in greater detail in Capriati (2024, ch. 2).

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

Therefore, in asking whether a machine can be considered rational, it is appropriate to focus not on internal (and unobservable) criteria but on what can be observed, thus providing an empirical standard and concluding that a machine is rational if it behaves rationally. This "behaviourist" definition of rationality is ultimately the one that matters to those who grant legitimacy to a decision-maker on the condition that they act rationally. In other words, what concerns such subjects is the decision-maker's behaviour, not the internal processes that determine it.

Russell and Norvig (2016, 40), in their famous "Artificial Intelligence: A Modern Approach", define a rational agent:

> (RA1) For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has.

This definition of rationality corresponds to instrumental rationality. It has the merit of translating the concept of rationality into that of performance (or behaviour): an agent is rational if it maximizes performance in its actions. Rationality is therefore not understood as a particular disposition of the mind (or thought), but as something that can be observed by examining the behaviour of a given agent. (RA1) is not the only definition that Russell and Norvig provide for a rational agent:

(RA2) A rational agent is one that acts so as to achieve the best outcome or, when there is uncertainty, the best expected outcome (Russell and Norvig, 2016, 4).

(RA3) A rational agent is one that does the right thing (Russell and Norvig, 2016, 39).

With reference to (RA3), it is necessary to clarify what it means to do the right thing. Among the main tasks of moral philosophy is to define what is meant by "the right thing". According to Russell and Norvig (2016, 39), in the world of AI, a consequentialist approach is commonly adopted, which

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

evaluates the behaviour of an agent based on the consequences that such behaviour produces. To evaluate the consequences of a certain behaviour, it is necessary to adopt a clear parameter of reference. In other words, assessing whether the consequences produced by agent A are better than those produced by agent B requires specifying the criteria according to which such a judgment is made – better in relation to a defined standard or value system.

This definition makes sense if a well-defined teleological perspective is introduced: if the expected output can be clearly defined, the more rational the agent will be, the more its behaviour allows it to approach that output.

Russell and Norvig (2016, 39) then proceed to distinguish between human rationality and machine rationality:

> Humans have desires and preferences of their own, so the notion of rationality as applied to humans has to do with their success in choosing actions that produce sequences of environment states that are desirable from their point of view. Machines, on the other hand, do not have desires and preferences of their own; the performance measure is, initially at least, in the mind of the designer of the machine, or in the mind of the users the machine is designed for.

This distinction does not deserve to be explored further. Regardless of whether it is a human being or a machine, an agent will be considered rational if, based on the objectives it intends to pursue, it acts to maximize the chances of achieving those objectives. At this moment, it is irrelevant to understand how these objectives are determined.

In conclusion and answering the question with which I opened the paragraph, the rationality of a machine can be assumed just as the rationality of any other living can be assumed.

2.4 All the Terms of the Issue

The terms of the issue are now clearer. The thesis “some conceptions of government are based on assumptions of rationality and, depending on the subject whose (greater) rationality is assumed, different conceptions of government arise" consists of the following elements:

(1) Some subjects (2) assume (3) the rationality of (4) certain others – and (5) machines can, in principle, be candidates to be such agents – and, based on these assumptions, (6) different conceptions of government arise.

(1) The subjects who confer legitimacy on a government, namely, the governed.

(2) The assumptions of rationality consist of the belief that certain subjects are rational.

(3) I refer to an instrumental conception of rationality, which considers rational those subjects who make decisions adequate to achieve certain ends.

(4) The subjects whose rationality is assumed are those who are assumed to be able to make political decisions adequate to achieve a certain end.

(5) Among the subjects whose rationality can be assumed, there are also the machines.

(6) Each conception of government justifies a model of government based on specific reasons.

In the following pages, I will aim to demonstrate how assumptions of rationality determine political legitimacy and to present a brief taxonomy of conceptions of government based on the subjects who are assumed to be (most) rational.

3. Assumptions of Rationality and Political Legitimacy

In this section, I will focus on the relationship between assumptions of rationality and political legitimacy – that is, I will aim to demonstrate (T): how assumptions of rationality politically legitimize a subject to make decisions.

We must, in dealing with (T), understand what conditions make it necessary to consider assumptions of rationality to legitimate a government.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

My claim is that it is necessary to assume that governments are rational only insofar as the political system is understood in epistemic terms, that is, insofar as there is the belief that:

(a) Some decisions are better than others;

(b) There exists a standard of correctness for decisions that is independent of the government's decision-making process;

(c) The government is considered the agent capable of discovering what these correct decisions are;

(d) The government is justified based on (c).

I will refer to this way of understanding political systems as the "epistemic conception" (Capriati, 2024). The reason why, in this case, it is necessary to assume the government's rationality is that, in such a political system conception, the government is politically justified in acting only if it is capable of making the right decision (admitting, beforehand, that there is a right decision and that it is independent of the government's decision-making procedure). In this sense, in order to make the right decision, the government must be rational – that is, it must adopt the appropriate means to achieve predetermined ends.

The connection to justification is straightforward: as stated in (d), the justification of the government depends on its capacity to discover what the correct decisions are. The government that acts rationally – and is therefore capable of discovering the correct decision – is the government considered justified and, consequently, the legitimate government.

Let us assume, for example, that there is a government that makes decisions by reading the coffee grounds at the bottom of a cup. Adopting an epistemic conception of the political system, could such a government be justified? Yes – but only if we were willing to assume (and demonstrate) that coffee provides rationally relevant information with respect to the decisions to be made. Therefore, regardless of the model of government, in a political system based on an epistemic conception, the government's action is considered justified if it is assumed that it acts rationally.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

To justify, therefore, the government of a machine, in a political system conceived in epistemic terms, it will be necessary to find that there is a widespread belief that the machine acts rationally.

3.1 Taxonomy of Government Conceptions Based on Rationality Assumptions

In some conceptions of government, the rationality of the subject legitimate to decide is assumed. What I will try to demonstrate is that, depending on the subject whose rationality is assumed, different conceptions of government arise (T_1). This means that the assumption of rationality is an element that determines the conception of government.

I will examine three conceptions of government, showing how, behind each of them, the subject whose rationality is assumed is different. In the case of the classical – or aggregative – democratic conception, the assumed rational subjects are individual citizens, whereas in the case of the deliberative democratic conception, rationality would be the prerogative of the group, understood as a whole. Finally, it will be shown how the possible delegation of decision-making power to a machine is necessarily linked to assuming the rationality of the machine itself.

Before presenting this taxonomy, as we have already seen, it is necessary to recall that by "conceptions of government", I do not refer to the array of norms and bodies that constitute the institutional apparatus of the state, but to the normative structure adopted to make decisions of public interest.

3.1.1 Aggregative Democracy

When we assume the rationality of individuals, a specific conception of government emerges. The conception of government that assumes individuals are rational, and therefore legitimizes each citizen to participate in decision-making, is democracy in its aggregative form – or the economic theory of democracy.

Jon Elster (1986) states that aggregative theories of democracy see the political process as a means rather than an end in itself. For these theories, the decisive political act would be a private rather than a public action, namely the individual and secret vote. These theories are thus united by a merely "aggregative" view of the democratic process: the aggregation of individual preferences or interests is the way to achieve a collective social choice. The aggregative version also considers citizens' preferences as fixed and predetermined. As Elster (1986, 128) says: "It is a market theory of politics, in the sense that the act of voting is a private act similar to that of buying and selling".

My claim is that, according to this conception of government, those decisions made by aggregating the individual preferences of subjects – whose rationality is assumed – are legitimate. The emblem of this conception of government is the electoral process. Through elections, each subject, equally rational, has equal opportunities to influence the decision.

An objection that may be raised is that, in such a conception of government, it is not necessarily assumed that individuals are rational, but simply that (1) they pursue their own personal interests, and (2) they are legitimate to make decisions because they are the ones who know their interests best. In this sense, individuals' personal interests are not necessarily rational. However, this objection can be easily overcome. In this context, in fact, the rationality of personal interests is not at issue. Rationality, understood in an instrumental sense, concerns the adequacy of means to ends – which coincide with personal interests – and therefore the ability of individuals to pursue their own interests better than anyone else (or, if necessary, to delegate their expression better than anyone else).

It is clear that in such a system, it is not expected that every decision will always be made by all of the individuals. In practice, there is a division of decision-making according to a principle of competence. Is this mechanism in contradiction with the idea that all individuals are rational? No, since the rationality of each individual is assumed, they are precisely rational enough to recognize they might not be able to decide on everything, realizing that certain decisions would be better if delegated.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

In such a conception, therefore, the assumption of rationality concerns individual subjects. The dimension is always collective and is built upon the sum of those individuals.

3.1.2 Deliberative Democracy

From a conceptual point of view, deliberative democracy aims to overcome the aggregative conceptions of democracy (Bächtiger, Dryzek, Mansbridge and Warren, 2018). It can be characterised as that set of conceptions in which public deliberation by free and equal citizens constitutes the heart of the political legitimacy of decision-making and self-governance processes (Bohman, 1998). The distinction between aggregative and deliberative theories is based on the focus of interest: while the former insists on the aggregation of individual preferences – and see voting as the most emblematic representation –, deliberative theories emphasize the transformative possibilities of preferences through dialogical and discursive processes. In the case of deliberative democracy, the subject whose rationality is assumed is the group of individuals. According to this conception of government, rationality is understood as the prerogative of the group rather than of single individuals.

Deliberative democracy has embraced and processed some of the criticisms directed at the idea that the individual is a rational subject. These criticisms are primarily driven by some psychological studies: as seen, these studies have questioned the hypothesis of perfect rationality of the individual (Kahneman, 1994; Stich, 1990). These studies have been accompanied by another literature – also of psychological origin – which argues that certain behaviours, if examined at the individual level, appear dysfunctional, while they may appear functional from a collective point of view (Mercier and Sperber, 2011, 2017; Mercier, 2020). Sperber and Mercier construct an argumentative theory that fits well with some of the main theses of deliberative democracy. More precisely, I refer to the idea that the human tendency to prefer evidence that confirms pre-existing beliefs and to ignore

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

those that contradict them derives from the ability to construct the best possible argument in support of a particular thesis. This tendency, then, would suggest that human reasoning abilities have evolved to persuade others and take positions in debates, rather than to seek the truth. In this sense, the confirmation bias would be evolutionarily more effective than impartial and disinterested truth-seeking.

The myth of the rational individual is the polemical target not only of psychological studies. Habermas – considered among the founding fathers of deliberative democracy argues –, consistently with the work of Sperber and Mercier, that rationality is a widespread phenomenon and not a quality exclusive to individuals. Moreover, Habermas attempts to overcome the limits of instrumental rationality by proposing a new type of rationality that emerges in a dialogical context and gives rise to the ideal discursive situation (Habermas, 1987). He calls this rationality "communicative" and immediately clarifies that it is not a subjective faculty (Habermas, 2015), but that it takes shape in dialogical and discursive processes.

According to a deliberative conception of democracy, therefore, the subject whose rationality must be assumed is not the individual nor each individual, but the group as a whole and the institutionalised discursive procedure they adopt to reach decisions. Individuals, whose behaviour, if evaluated in itself, can easily be considered irrational, in interacting with each other, would select courses of action informed by rationality.

3.1.3 Machine-Government

Assuming that the group, through interactions among individuals, acts rationally is a fact far from indisputable. Common sense, as well as other psychological studies, seems to go in the opposite direction: in groups, dynamics often arise that negatively affect decisions. Sunstein (2002, 187), for example, observed how deliberation can generate even worse decisions because the "law of polarization" prevails in groups, leading people to cling to predetermined positions.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

Questioning the rationality of the group prepares the ground for assuming the rationality of another subject: the machine. I have already anticipated, in paragraph 1.4, the issues related to the possibility of assuming the rationality of an artificial agent. There are no particular reasons preventing the assumption of the rationality of a machine.

No expression explicitly refers to this: even the term "Algocracy" (Danaher, 2016), as it has been defined, does not seem to fit our case.[7] We could name the conception of government in which a machine is authorised to make decisions of collective interest "machine-government".

A machine-government is a system in which machines make decisions of collective interest. The hypothesis I am proposing is not that of a machine dictatorship. Rather, "deciding" in this context refers to the mechanism by which inputs coming from human subjects are selected and transformed. In other words, the role of the machine is to collect, weigh, and process individual preferences, with the aim of producing a decision whose justification lies in the idea that the machine assigned to this task is rational and that, more generally, a machine can be a rational entity, or certainly more rational than other agents that could decide.[8]

Some doubts remain regarding the exact determination of the machine-government. I will now present a concrete example.

Delegating public interest decisions to a machine is still a distant reality. However, the Habermas Machine (hereafter HM) (Tessler et al., 2024)

[7] "I use it to describe a particular kind of governance system, one which is organised and structured on the basis of computer-programmed algorithms. To be more precise, I use it to describe a system in which algorithms are used to collect, collate and organise the data upon which decisions are typically made and to assist in how that data is processed and communicated through the relevant governance system" (Danaher, 2016, 247). In the model I have in mind, the machine does not merely organize the data on which decisions are based, but autonomously makes the decisions itself.

[8] Rationality cannot be understood as a necessary and sufficient condition for legitimizing a delegation of decision-making power, since it is clear that when assessing the quality of a political decision, one cannot rely solely on the criterion of rationality. A political decision, in fact, to be considered legitimate, requires additional conditions, such as epistemic or ethical ones. Examples may include transparency or the possibility for users' verification. This analysis, however, explicitly focuses on the criterion of rationality: it is on the basis of this criterion, in fact, that the taxonomy I have just presented is constructed.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

represents a concrete realization of a system capable of collecting, processing, and transforming preferences. HM is a project by Google DeepMind based on a system of LLM – large language model (such as Chat GPT). The name "HM" is a tribute to Habermas's theory of communicative action, according to which when rational subjects deliberate in an ideal discursive situation, they manage to reach an agreement.

HM was designed to improve collective decision-making processes in various fields. For example, it can be used for contract negotiations, conflict resolutions, political discussions, and citizens' assembly (Tessler et al., 2024, 1).

In these areas, HM acts as a "caucus" mediator. A caucus mediator is defined as one who privately meets each interlocutor before formulating a proposal that can be collectively accepted (Moore, 1987). HM does not merely mediate the discussion but also formulates decisions that are then submitted for approval by the group members. In this sense, the machine acts as a processor and transformer of various instances to shape a decision that best meets the needs of the greatest number of subjects.

4. Conclusions

What is the relationship between rationality assumptions and the legitimacy of political decision-makers? Is it possible to legitimate a machine to make decisions of collective interest?

To answer these questions, I began by illustrating the problem:

Can conceptions of government be distinguished according to the subject whose rationality is assumed?

Firstly, I clarified the elements that compose this question:

- What are the assumptions of rationality, and what purpose do they serve?
- What is a conception of government?
- What is rationality?

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

- Can the rationality of a machine be assumed?

Through these questions, it emerged that five elements must be taken into consideration:

(1) The subjects who confer legitimacy on a government;

(2) The assumptions of rationality;

(3) Rationality;

(4) The decision-making subjects;

(5) Machines as decision-making subjects;

(6) The conceptions of government.

After having presented all the elements that constitute my research question, I focused on how (and why) assumptions of rationality politically legitimize a subject to decide. Some conceptions of government are based on rationality, that is, on the idea that the decision-maker must inform their action by rationality. These are the conceptions of government that operate within a political system based on the epistemic account. According to these conceptions, therefore, the subject legitimate to decide is the one who decides rationally. Depending on the legitimate subject who decides – that is, the one considered rational – different conceptions of government arise.

More precisely, I examined three conceptions of government (and corresponding rationality assumptions):

(A_1) aggregative democracy (which assumes the rationality of individuals);

(B_1) deliberative democracy (which assumes the rationality of the group as a whole);

(C_1) machine-government (which assumes the rationality of the machine).

In summary, (T) in some conceptions of government, assuming the rationality of the decision-making subject is a necessary requirement to consider such government legitimate, and (T_1) the assumptions of rationality are central in determining the nature of the conception of government.

In this paper, I have addressed the fictional nature – fictional because unsupported by facts – of the political decision-maker. Returning to the

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

question that opened this contribution: should we entrust the government of public affairs to a machine? If one were to start from an epistemic conception of political systems, the answer could be affirmative, provided that the machine is assumed to be rational. Whether the machine – just like political decision-makers – is actually rational is another matter, one that opens up new avenues for empirical research.

References

Bächtiger A., Dryzek J. S., Mansbridge J. and Warren M. (2018). Deliberative Democracy: An Introduction, in A. Bächtiger, J. S. Dryzek, J. Mansbridge and M. Warren (eds.), *The Oxford Handbook of Deliberative Democracy* (Oxford University Press), 1-31.

Beetham D. (1991). *The Legitimation of Power* (Palgrave).

Bohman J. (1998). Survey article: The coming of age of deliberative democracy, in *Journal of Political Philosophy*, n. 6(4), 400-425.

Capriati P. (2024). *Macchine che decidono: prospettive di automazione dei processi decisionali in contesti democratici*, PhD thesis, University of Palermo.

Commoner B. (1965). Biochemical, Biological and Atmospheric Evolution, in *Proceedings of the National Academy of Science*, n. 53, 1183-1194.

Danaher J. (2016). The threat of algocracy: Reality, resistance and accommodation, in *Philosophy & Technology*, n. 29(3), 245-268.

Elster J. (1986). The Market and The Forum: Three Varieties of Political Theory, in J. Elster and A. Hylland (eds.), *Foundations of Social Choice Theory* (Cambridge University Press), 103-132.

Habermas J. (1987). *The Theory of Communicative Action* (Vol. 1) (Beacon Press).

Habermas J. (2015). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (John Wiley & Sons).

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

Kahneman D. (1994). New Challenges to the Rationality Assumption, in *Journal of Institutional and Theoretical Economics*.

Kahneman D. (2011). *Thinking, Fast and Slow* (Penguin).

Kolodny N. and Brunero J. (2013). Instrumental rationality, in *Stanford Encyclopedia of Philosophy*.

Maturana H. R. and Varela F. J. (2012). *Autopoiesis and Cognition: The Realization of the Living* (Vol. 42) (Springer Science & Business Media).

Mercier H. (2020). *Not Born Yesterday: The Science of Who We Trust and What We Believe* (Princeton University Press).

Mercier H. and Sperber D. (2011). Why Do Humans Reason? Arguments for an Argumentative Theory, in *Behavioral and Brain Sciences*, n. 34.

Mercier H. and Sperber D. (2017). *The Enigma of Reason* (Harvard University Press).

Moore C. W. (1987). The caucus: Private meetings that promote settlement, in *Mediation Q.*, n. 87.

Penrose R. (1991). The emperor's new mind, in *RSA Journal*, n. 139(5420), pp. 506-514.

Russell S. J. and Norvig P. (2016). *Artificial Intelligence: A Modern Approach* (Pearson).

Smithers T. and Moreno A. (1994) (Eds). Notes for the Workshop on the Role of Dynamics and Representation in Adaptive Behaviour and Cognition, 9 and 10 December, Palacio de Miramar, San Sebastian (Spain).

Smithers T. (1997). Autonomy in Robots and Other Agents, in *Brain and Cognition*, n. 34(1), pp. 88-106.

Steels L. (1995). When are Robots Intelligent Autonomous Agents?, in *Robotics and Autonomous Systems*, n. 15(1-2), pp. 3-9.

Stich S. P. (1990). *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation* (MIT Press).

Sunstein C. R. (2002). The Law of Group Polarization, in *The Journal of Political Philosophy*, n. 10 (2), pp. 175-195.

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550

Tessler M. H., Bakker M. A., Jarrett D., Sheahan H., Chadwick M. J., Koster R., ... and Summerfield C. (2024). AI can help humans find common ground in democratic deliberation, in *Science*, n. 386(6719), eadq2852.

Thaler R., and Sunstein C. R. (2008). *Nudge: Improving Decisions about Health, Wealth and Happiness* (Yale University Press).

Turing A. M. (1950). Computing Machinery and Intelligence, in *Mind*, n. 59, pp. 433-460.

Weber M. (1964). *The Theory of Social and Economic Organization* (Simon and Schuster).

Weber M. (1978). *Economy and society: An outline of interpretive sociology* (University of California Press).

ISSN 2724-6299 (Online)
https://doi.org/10.60923/issn.2724-6299/22550