

## UNDERSTANDING RATIONALITY IN HOBBS AND HUME

HUN CHUNG, Department of Philosophy, Rochester Institute of Technology, Rochester, NY, USA

CHUNG, H.: Understanding Rationality in Hobbes and Hume  
FILOZOFIA 69, 2014, No 8, pp. 687-696

Many commentators think that Hobbes was committed to an instrumental view of rationality which foreshadows that of David Hume. The Humean conception of instrumental rationality is a conjunction of the following two claims: (a) no preferences or desires can properly be said to be irrational in themselves, and (b) the role of reason or rationality can only be confined to informing the agent with true beliefs about the world, and revealing the most effective means that could satisfy the agent's current ends whatever they happen to be. In this paper, I argue that, unlike what many people think, a careful reading of Hobbes shows that he was committed to neither of these claims.

**Keywords:** Hobbes – Hume – Instrumental rationality – Practical rationality – The Humean view of rationality

**1. David Hume's Conception of Instrumental Rationality.** Many commentators interpret Hobbes as being committed to an instrumental view of rationality which foreshadows that of David Hume. It is somewhat controversial whether Hume had actually adopted such view of rationality.<sup>1</sup> However, it is quite common to attribute such instrumental conception of rationality to Hume, and, regardless of whether Hume had actually held this view or not, I would like to show that at least Hobbes wasn't committed to such view. Of course, the view that Hobbes was not committed to a purely instrumental view of rationality is not entirely new.<sup>2</sup> However, in this paper, I will provide new interpretations of some of the key textual evidence of Hobbes (which has been relatively neglected by scholars) to substantiate my particular interpretation.

It is best to understand what is known as the "Humean conception of practical rationality (i.e. instrumental rationality)" by looking at the passages that Hume wrote himself. According to Hume,

We speak not strictly and philosophically when we talk of the combat of passion and of reason. Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them. [Hume 1984: *A Treatise of*

<sup>1</sup> Nick Sturgeon's unpublished manuscript, "Hume on Reason and Passion", deals with this issue with careful detail in depth.

<sup>2</sup> See, for instance, Bernard Gert, "Introduction" contained in *Thomas Hobbes – Man and Citizen*, Hackett, 1991.

*Human Nature*, Book II, Part III, Section III: p. 462]<sup>3</sup>

‘Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. ‘Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me. ‘Tis as little contrary to reason to prefer even my own acknowledge’d lesser good to my greater, and have a more ardent affection for the former than the latter. [Hume 1984: *A Treatise of Human Nature*, Book II, Part III, Section III: p. 463].

The thesis that Hume is interpreted as advancing in the first quoted passage is, what is generally known as, the “No-Combat Thesis”; namely, the thesis that there is no way for one’s desires or preferences to contradict the dictates of reason and rationality. In other words, according to Hume, it is not the role of reason or one’s other rational faculties to evaluate and oversee whether one’s desires or preferences are rational or reasonable. One’s desires and preferences are simply things that the individual just happens to have whose contents over which the individual’s reason or rationality has no control.

The result of this, according to Hume, is that there cannot be any preferences or desires that can properly be called *irrational*, and, thereby, be seen as in conflict with one’s rationality. According to Hume, *no preferences are irrational* -- it is not irrational to prefer the destruction of the whole world to the scratching of my finger; it is not irrational to prefer my total ruin to the least amount of uneasiness that a totally unrelated foreigner might feel; and it is not even irrational to prefer what I believe to be worse for me to what I believe to be better for me -- and it is not the job of one’s reason and rational faculties to encourage one to form rational preferences – because, strictly speaking, preferences *can neither be rational nor irrational*.

Then, according to Hume, what are the roles of reason and rationality in the process of deriving one’s preferences? According to Hume, there are two major roles that reason and rationality play in this area: one is to help the agent form true beliefs about the world, and the other is to help the agent choose the most effective means to satisfy the agent’s current preferences or desires (whatever they happen to be) in the light of these true beliefs.

This means that there are two basic ways that one’s preferences or desires might be (rather imprecisely) called irrational or unreasonable. One way is for one’s preferences to be based on a *false belief*; such as when one prefers not to go to the bathroom at the middle of the night because one falsely believes that there is a scary monster living in the bathroom that only shows up at night. Another way is for one to prefer to use a specific means for a given end which is highly *ineffective* in accomplishing the end that one has in mind; such as when one prefers to use a toothbrush to open a bottle of beer.

‘tis only in two senses, that any affection can be call’d unreasonable. First, When a passion, such as hope or fear, grief or joy, despair or security, is founded on the sup-

---

<sup>3</sup> The page numbers refer to the Penguin (1984) edition.

position or the existence of objects which really do not exist. Secondly, When in exerting any passion in action, we chuse means insufficient for the design'd end, and deceive ourselves in our judgment of causes and effects. [Hume 1984: *A Treatise of Human Nature*, Book II, Part III, Section III, p. 463].

However, even though we commonly call these two types of preferences as instances of irrational or unreasonable preferences in our ordinary life, according to Hume, it is not, strictly speaking, the preferences themselves that are irrational or unreasonable, but only the accompanying beliefs that are so.

In short, a passion must be accompany'd with some false judgment, in order to its being unreasonable; and even then 'tis not the passion, properly speaking, which is unreasonable, but the judgment. [Hume 1984: *A Treatise of Human Nature*, Book II, Part III, Section III, p. 463]

So, according to the Humean conception of practical rationality, it might be irrational for Jill, who is currently living a moderately decent life, to prefer to commit suicide by swallowing a chewing gum. But, the reason why we can (rather imprecisely) call such preference irrational is not because preferring to commit suicide when one is living a moderately decent life is itself irrational, but rather, because Jill has not chosen an effective means to achieve her given end, which is to kill herself. If Jill had preferred to kill herself and end her life by using a very effective means (e.g. jump off a 40-story building), then Hume would say that there is nothing irrational about her preferences. In other words, according to the Humean conception of practical rationality, no preferences (not even a preference to kill one's life without having any good reason to do so) can, strictly speaking, be irrational or unreasonable in itself.

Let's summarize the Humean conception of practical rationality into the following formula:

**[Humean Conception of Practical Rationality (i.e. Instrumental Rationality)]**

: The Humean conception of practical rationality (i.e. the purely instrumental view of rationality) is the conjunction of the following two claims:

- (a) No preferences or desires, by themselves, are irrational or rational.
- (b) (The Role of Reason and other Rational Faculties): the basic role of reason and other rational faculties are confined to the following two roles;
  - (b-1) To inform the agent with *true beliefs* about the world, and
  - (b-2) To inform the agent with *the most effective means* to achieve a given end in the light of these true beliefs about the world.

\* That is, reason and rationality say nothing about the agent's ends themselves.

**2. Was Hobbes Committed to the Humean Conception of Instrumental Rationality?** Now, the question is: "Did Hobbes endorse the Humean conception of instrumental rationality as formulated above?" Many people think that he did. Consider Jean Hampton's early interpretation:

So, for Hobbes, (...) Rationality would therefore be regarded by him as having instrumental value; a rational man would be one whose reason would serve his desires well by determining correctly how those desires could be satisfied. (...) Moreover, (...), Hobbes makes his instrumentalist notion of reason crystal clear (...) [Hampton 1986<sup>4</sup>, p. 35]

David Gauthier's interpretation of Hobbes is similar.

If we accept also Hobbes' contention that 'those actions are most reasonable, that conduce most to their ends' (E.W. iii, p. 133), we can then derive from 'a man must do X to secure what he wants', the further conclusion 'a man, if rational, will do X.' (...) In the state of nature *reason is the slave of the passions*; hence to refer to a right to do what one naturally endeavours to do is otiose. [Gauthier 1969<sup>5</sup>, p. 21, 35, emphasis added]

Hobbes certainly did not deny that one's reason and rational faculties can play such an instrumental role in one's deliberation process. In fact, one of the major reasons why Hobbes advises one to bring one's prudence, reasoning, and scientific knowledge to bear on one's deliberation process is to enable the agent to determine his/her final preferences on the basis of "the greatest and surest prospects of consequences."

... so that he who hath by *experience* or *reason* the greatest and surest prospect of consequences deliberates best himself, and is able, when he will, to *give the best counsel unto others*. [Hobbes 1994: *Leviathan*<sup>6</sup>, Chapter VI, Paragraph 57]

And, here, we might say that the reason why Hobbes advises one to combine "the greatest and surest prospects of consequences" with one's deliberation process is because, having such information will, generally, reduce the chances of one's final preferences to be based on *false beliefs*, and such information will also, very likely, reveal what *the most effective means* or *what the most effective course of action* could be to achieve the end that one has in mind. In other words, Hobbes does think that one's reason and rational faculties can perform the specific roles that are described in clause (b) of the formula of "the Humean conception of practical rationality" stated above.

However, for Hobbes, the roles that reason and rationality play are *not confined* to the two activities of preventing the formation of false beliefs and revealing the most effective means to a given end. This is because Hobbes thought that any preferences the satisfaction of which are inconsistent with the achievement of one's long-term self-preservation are irrational. Consider the following passages:

---

<sup>4</sup> Jean Hampton: *Hobbes and The Social Contract Tradition*. Cambridge University Press 1986.

<sup>5</sup> David Gauthier: *The Logic of Leviathan*. Oxford University Press 1969..

<sup>6</sup> All quotations from Hobbes's *Leviathan* are from the version edited by Edwin Curley (Hackett, 1994)

Moreover, *the greatest of goods for each is his own preservation*. For nature is so arranged that all desire good for themselves. Insofar as it is within their capacities, *it is necessary to desire life, health*, and further, insofar as it can be done, security of future time. [Hobbes 1991: *De Homine*,<sup>7</sup> Chapter XI, Section 6, emphasis added]

Here, we can see that Hobbes regards *self-preservation* as the greatest good for each and every individual. Now, consider Hobbes's definition of *madness*:

In sum, all passions that produce strange and unusual behavior are called by the general name of madness. [Hobbes 1994, *Leviathan*: Chapter VIII, Paragraph 20]

Here, we can see that Hobbes defines madness as any passion that produces strange and unusual behavior. Then, what kind of behavior did Hobbes regard as strange and unusual? Hobbes was not explicit on this matter, but we may infer what Hobbes thought to be strange and unusual behavior by considering what kind of behavior Hobbes thought to be *normal*. Here, we should remind ourselves from the previous quoted passage that, according to Hobbes, it is *necessary to desire one's long-term self-preservation*. From this, we may say that, for Hobbes, any behavior that is consistent with one's long-term self-preservation is *normal*, and, conversely, any behavior that is harmful to securing one's long-term self-preservation is *strange* and *unusual*. If our inference so far is correct, then, for Hobbes, any passion that produces behavior that is detrimental to securing one's long-term self-preservation would be madness. We can further confirm that this conjecture is correct by looking at what Hobbes thought to be the two representative cases of madness:

The passion whose violence or continuance maketh madness is either great vain-glory, which is commonly called *pride* and *self-conceit*, or great dejection of mind. [Hobbes 1994, *Leviathan*: Chapter VIII, Section 18]

Both vain-glory and great dejection are harmful to one's long-term self-preservation; the former, making people too aggressive, and the latter, making people self-destructive. Hobbes thought that being overwhelmed by either of these two passions is what makes a person *mad*. So, when Hobbes claims that it is necessary to desire one's long-term self-preservation, we should remind ourselves that there is actually a hidden clause; for Hobbes, it is necessary to desire one's long-term self-preservation *provided that one is not mad* – that is, provided that one is not suffering from any irrational passions.

On the face of it, this already seems to contradict clause (a) of the above formula – which states that no preferences or desires (or passions), by themselves, are irrational. However, we would need to verify that when Hobbes talks about irrational preferences or instances of madness, he is not merely talking about preferences that are based either on false beliefs about the world or on false beliefs about the effective means for a given end,

---

<sup>7</sup> All quotations from *De Homine* are from the version edited by Bernard Gert, *Thomas Hobbes – Man and Citizen*, (Hackett, 1991)

but is also talking about preferences, which, he thinks, are *irrational in themselves*. But, before we deal with clause (a), let's first start with clause (b).

**2-1. Hobbes Did Not Endorse Claim (b).** According to clause (b), the role of rationality and reason is confined to providing the agent with true beliefs and revealing what is the best means to achieve whatever end that the agent just happens to have. To put it in another way, clause (b) claims that the role of reason and rationality does not extend to directing one's desires and preferences towards a specific direction; that is, reason and rationality say nothing about the agent's ends *themselves*.

However, we can see that Hobbes did not endorse clause (b) by reminding ourselves of what Hobbes said about *moral science* and *the laws of nature*. According to Hobbes, the full culmination of one's reason or rationality is scientific knowledge.

For REASON, in this sense, is nothing but reckoning (that is, adding and subtracting) of the consequences of general names agreed upon for the marking and signifying of our thoughts... [Hobbes 1994, *Leviathan*: Chapter V, Section 2].

...till we come to a knowledge of all the consequences of names appertaining to the subject in hand; and that is it men call SCIENCE. [Hobbes 1994, *Leviathan*: Chapter V, Section 17].

When the matter is about what is good for individuals and societies in general, the scientific field that provides answers to these questions is what Hobbes calls *moral science* or *moral philosophy*.

For moral philosophy is nothing else but the science of what is good and evil in the conversation and society of mankind. [Hobbes 1994, *Leviathan*, Chapter XV, Section 40].

Then, what are the specific contents of what Hobbes calls "moral science" or "moral philosophy?" They are what Hobbes calls *the laws of nature*:

A LAW OF NATURE (*lex naturalis*) is a precept or general rule, found out by reason, by which a man is forbidden to do that which is destructive of his life or taketh away the means of preserving the same, and to omit that by which he thinketh it may be best preserved. [Hobbes 1994, *Leviathan*: Chapter XIV, Section 3].

And the science of them [the laws of nature] is the true and only moral philosophy. [Hobbes 1994, *Leviathan*, Chapter XV, Section 40].

In other words, the laws of nature, according to Hobbes, are prescriptive rules discovered by reason that provide people with practical guidelines on how to secure one's own self-preservation.

So, here is where we are so far. According to Hobbes, a person's reason or rationality, if developed fully, results in scientific knowledge. One branch of science, which concerns what is good for individuals and societies, is moral science or moral philosophy. The specific contents of moral science or moral philosophy are the laws of nature. But,

the laws of nature are none other than general rules or precepts that give specific guidelines to secure one's long-term self-preservation. Here, we see that Hobbes actually thought that reason and rationality should properly *guide* and *direct* the agent's *ends* to aim for his/her long-term self-preservation.

This shows that Hobbes did not accept clause (b) of the formula of the Humean conception of instrumental rationality – which, in effect, claims that reason and rationality can say nothing about the agent's ends themselves. According to Hobbes, reason or rationality does say what one's basic end should be; and that is to achieve one's long-term self-preservation.

**2-2. Hobbes Did Not Endorse Claim (a).** To this, one might object that the fact that the laws of nature aim for self-preservation and that these laws are discovered by right reason does not, strictly speaking, imply that it is the proper role of reason and rationality to direct one's ends. Rather, “the laws of nature” is a perfect example that shows how the basic role of reason and rationality is confined to the instrumental role described in clause (b). That is, according to this interpretation, the laws of nature, which are discovered by right reason, merely tell people *the most effective means* to achieve what they most basically *want*: namely, their own self-preservation.

To take this approach, one has to assume that Hobbes thinks that it is *a matter of fact* that *everybody* has a basic desire for self-preservation which is stronger than any other desire they happen to have and that the sole job of reason is to choose the most effective means to achieve this end which is already given. This seems to be the view adopted by those who attribute the Humean instrumentalist conception of rationality to Hobbes.

If we accept Hobbes's view that man is a self-maintaining engine then ... Men want, and necessarily want, to preserve themselves. Therefore, whatever can be shown to be a condition of human preservation, is thereby shown to be a means to man's end. From premises of the form ‘X is a necessary means to self-preservation’, Hobbes can derive conclusions of the form ‘a man must do X to secure what he want’. ... we can then derive from ‘a man must do X to secure what he wants’, the further conclusion ‘a man, if rational, will do X’. [Gauthier 1969, p. 21].

Here, we could specify the variable X by “the laws of nature”. According to Gauthier's interpretation, Hobbes thinks that we are “self-maintaining engines” who just cannot help but to desire our own self-preservation, and the main role of reason and rationality is to identify the most effective means – which is to follow the prescriptions of the laws of nature – to secure this already given end. So, when Hobbes seems to be advancing a normative claim that “one should follow the prescriptions of the laws of nature”, he is, strictly speaking, not recommending one to desire self-preservation as a specific end, but rather is only recommending one to choose the necessary means to secure what one already wants the most.

If this is the case, then, what's so wrong with people who are overwhelmed by the passion for vain-glory and seek glory at the expense of their own self-preservation, who,

according to Hobbes, are *irrational* and *mad*? Isn't Hobbes, here, criticizing the specific ends to which glory-seeking people are committed? According to Hampton, what Hobbes thinks is wrong with glory-seeking people is not that they are pursuing the *wrong end*, but rather that they are seeking their own self-preservation *badly*.

So how can Hobbes have an instrumentalist conception of rationality when he is prepared to label as irrational those people who don't act to pursue their self-preservation? He can have such a conception if that label's meaning is roughly equivalent to 'imprudent,' that is, if the label is critical of such people not because they are pursuing an object rather than self-preservation but because they are perceived to be pursuing self-preservation badly. (...) Hobbes's condemnation would thus convict them of an error in their [instrumental] reasoning, not an error in what they were desiring. [Hampton 1986, p. 36].

This is incorrect. In order for Hampton's interpretation to be correct, she would have to show that Hobbes thinks that it is perfectly rational to seek glory (given that it is pursued effectively) if one truly desires glory and that it is OK to commit suicide (again, given that it is performed effectively) if one truly desires to conclude one's life. However, as we have seen, when Hobbes characterizes certain people as being mad or irrational, he is specifically criticizing the *abnormal passions* and *desires* that underlie their extreme behaviors. Recall how Hobbes defines "madness":

In sum, all *passions* that produce strange and unusual behavior are called by the general name of madness. [Hobbes 1994: *Leviathan*, Chapter VIII, Paragraph 20 emphasis added]

Again, that *madness* is nothing else but too much *appearing passion* ... [Hobbes 1994: *Leviathan*, Chapter VIII, Paragraph 23 emphasis added]

The opinions of the world, both in ancient and later ages, concerning *the cause of madness*, have been two. Some *deriving them from the passions*; some from demons or spirits [...] The former sort, therefore, called such men madmen; but the latter called them, [...] men possessed. [Hobbes 1994: *Leviathan*, Chapter VIII, Paragraph 24 emphasis added].

It is apparent that what Hobbes is criticizing here is the passions themselves. And it seems that it would be too much of a stretch to think of what Hobbes calls "madmen" as being merely *imprudent* in our ordinary sense of the term as Hampton suggests. As we have already seen previously, somebody might happen to have an extremely strong desire for glory, and this person might very intelligently choose the most effective means to satisfy this basic desire. This person is certainly *not imprudent* in the sense suggested by Hampton; nonetheless, the person will still count as mad by Hobbes's own standards, as we can quite clearly see in the following passage:

For what argument of madness can there be greater than to clamour, strike, and throw stones at our best friends? Yet this is somewhat less than such a multitude will



do. For they will clamour, fight against, and destroy those by whom all their lifetime before they have been protected and secured from injury. And if this be madness in the multitude, it is the same in every particular man. [Hobbes 1994: *Leviathan*, Chapter VIII, Paragraph 21].

So, when Hobbes is labeling glory-seekers as “mad” or “irrational”, what he is really criticizing is not merely that these people are choosing *ineffective means* to secure their own self-preservation, but also that these people are being overwhelmed by *the wrong kind of desires* – namely, the desire for (vain) glory – and, thereby, not desiring their own self-preservation strongly enough.

This shows that Hobbes did not endorse clause (a) of the above formula of the Humean conception of practical rationality – which claims that no desire or preferences are, by themselves, irrational. What Hobbes is criticizing about glory-seekers is their *ends themselves*, not just the means.

Moreover, since it is clear that Hobbes acknowledges the existence of mad people who do not desire their own self-preservation strongly enough, it is not entirely correct to interpret Hobbes as regarding human beings as what Gauthier calls “self-maintaining engines”, since this implies that everybody, *as a matter of fact*, has a basic desire for self-preservation which is *stronger than* any other desire they have or would have.

Not only does Hobbes think that it is possible for somebody to prefer glory over self-preservation, he also thinks that there is a certain proportion of the human population who actually do prefer glory over self-preservation in the state of nature where government authority is absent.

In the state of nature there is in all men a will to do harm, but not for the same reason or with equal culpability. One man practices the equality of nature (...) this is the mark of *modest man* (...) Another, supposing himself superior to other, wants to be allowed everything (...) that is the sign of an aggressive character. In his case, the will to do harm derives from *vainglory*. [Hobbes 1997, *On the Citizen*<sup>8</sup>: Chapter 1, Section 4 emphasis added].

Hobbes thinks that it is precisely this fact that makes it inevitable for the state of nature to dissolve into a state of war of all against all.

So, it seems that Bernard Gert’s interpretation of Hobbes’ theory of reason was generally going in the right direction when he claimed:

For Hobbes, reason provides a genuine guide to conduct, one applicable to all rational men; it is not merely a method whereby each man attempts to harmonize or maximize his particular passions. *That is, for Hobbes reason is not, or at least should not be, the slave of the passions, rather the passions are to be controlled by reason.* This is not to deny that “every man by reasoning, seeks out the means to the

---

<sup>8</sup> All quotations from *On the Citizen* are from the version edited by Richard Tuck, (Cambridge University Press, 1997).

end which he propounds to himself” but reason does more than this, it has an end of its own, avoidance of violent death. [Gert, “Introduction” in Hobbes 1991: *Man and Citizen*, p. 13 emphasis added].

The more surprising fact is that, in a later article, even Jean Hampton retracts her early interpretation of Hobbes and concedes that Hobbes did think that certain desires or preferences are irrational in themselves.

Elsewhere I argued that there was no inconsistency between these passages [i.e. passages where Hobbes apparently criticizes the preferences of glory-seeking people] and that theory of value [i.e. the desire(preference)-satisfaction theory of good]. *But now I have second thoughts*. [Hampton 1992,<sup>9</sup> p. 340 emphasis added].

... Hobbes does appear to criticize certain basic desires themselves, and not merely action from them, as irrational [Hampton 1992, p. 342].

In sum, based on numerous pieces of textual evidence as well as the reasons provided above, we can see that Hobbes did not endorse any one of the two clauses (a) and (b) in the formula of the Humean conception of practical rationality. That is, Hobbes did think that there are certain desires or preferences that are irrational in themselves. Moreover, Hobbes also thought that the role of reason and rationality is not and should not be confined to merely that of informing true beliefs about the world as well as the most efficient means to achieve a given end, but also to *guide the agent's end towards the right direction*. Therefore, we can conclude that Hobbes was not committed to a purely instrumental conception of rationality later presented by Hume.

## References

- GAUTHIER, D. (1969): *The Logic of Leviathan*. Oxford University Press.  
GERT, B. (1991): “Introduction” contained in *Thomas Hobbes – Man and Citizen*. Hackett.  
HAMPTON, J. (1986): *Hobbes and The Social Contract Tradition*. Cambridge University Press.  
HAMPTON, J. (1992): Hobbes and Ethical Naturalism. *Philosophical Perspectives*, 6. Ethics.  
HOBBES, T. (1991): *Man and Citizen*. (Edited by Bernard Gert.) Hackett.  
HOBBES, T. (1994): *Leviathan*. (Edited by Edwin Curley.) Hackett.  
HOBBES, T. (1997): *On the Citizen*. (Edited by Richard Tuck.) Cambridge University Press.  
HUME, D. (1984): *A Treatise of Human Nature*. Penguin.  
STURGEON, N.: Hume on Reason and Passion (unpublished manuscript).

---

Hun Chung  
Department of Philosophy  
Rochester Institute of Technology  
92 Lomb Memorial Drive  
Rochester, NY 14623-5604 USA  
e-mail: hunchung1980@gmail.com

---

<sup>9</sup> Jean Hampton, “Hobbes and Ethical Naturalism”, *Philosophical Perspectives*, 6. Ethics, 1992.