Are the Principles of Interpretation *A Priori*?

Naoyuki Shiono

Abstract

Interpretationism in the philosophy of mind assumes that certain basic principles of interpretation, collectively called the Principle of Charity, are constitutive of the practice of attribution of propositional attitudes, and should be accorded *a priori* status. This essay points out some serious misconceptions embodied in this assumption, and argues that the fundamental laws governing propositional attitudes should be conceived of as empirical rather than *a priori*.

I

Interpretationism is one of the major approaches in the philosophy of mind, and it is most notably associated with Davidson, who argues that propositional attitudes are attributed to human agents by the practice he calls "interpretation". An interpreter observes an agent, and tries to understand what he believes, what he wants, and what he does, as well as figure out the meanings of his utterances. This entire practice through which we come to understand an agent from scratch is "radical interpretation".

Propositional attitudes like beliefs and desires are not something we can directly observe, in the way we can perceptually be aware of the shape and colour of medium-sized physical objects. In order to know that an agent has, say, a belief that the wine bottle is empty, or a desire to go to the shopping mall, we usually need to make some inference from what is more immediately accessible to observation. Actions, on the other hand, may seem much closer to observation, because we sometimes say that we see someone raise his hands or open his bag. However, describing him as performing an intentional action presupposes that he has some relevant belief and desire, such as wanting to hail a taxi or believing that his notebook is in his bag. Therefore, judging that he has performed an intentional action involves more than simply watching his body move in a specific way. The meanings of what someone says, likewise, are not something we can directly hear. We usually assume that the agent speaks a familiar language and given this assumption the meanings of his utterances can be comprehended without difficulty. But the understanding of a language, according to interpretationism, is precisely what can only be achieved when we are in a position to

attribute to the agent appropriate propositional attitudes and actions.

These remarks do not imply that when we want to understand an agent, we start by noting the purely physical characteristics of what he does and says, describing his body in Newtonian terms as making a movement of certain velocity and direction, or emitting a sound sequence of a particular waveform. That, obviously, is not what actually happens. We generally describe him as performing an intentional action right away, spontaneously understand his utterances, and in some cases recognize his beliefs and desires quite transparently.

The point is rather that we can attribute propositional attitudes, actions and meanings to an agent only if they cohere with each other and with what is accessible to observation, and they as a whole exhibit a pattern, a structure characteristic of our understanding of someone as an agent. In other words, one attribution is legitimate only if we are ready to make other associated attributions, and they all contribute in making sense of what we observe about him. The idea of radical interpretation, where we try to make those attributions from the scratch, is perhaps nothing more than a thought experiment, but it is useful in laying out the kind of structure that must be manifested in any attempt at interpretation.

<center>II</center>

Proponents of interpretationism emphasize that rationality is the key concept in elucidating this requisite structure. They hold that any attempt at understanding an agent must respect the fundamental principle, which they call the Principle of Charity, that interpretation should be conducted in such a way as to maximize the rationality of the agent. It is this statement and some of its related assertions that this essay aims to examine. But first, we have to look more closely at what is involved.

When Quine first introduced the notion of the Principle of Charity, it was restricted to the modest requirement that the agent's utterances should not be translated in a way to make him committed to an obvious logical error, where a logical error is a contradiction in propositional and first-order predicate logic. Thus, when we find a native assent to p and also to bloop-p, a translation manual which tells us to translate "bloop" into *not* is a bad manual.[1] We are better advised to doubt the adequacy of the manual rather than to see the native as believing in contradictory propositions.

It was Davidson who made a much wider use of the Principle of Charity and applied it "across the board". There are a number of respects in which it was expanded, and Rescorla sorts them out into three components.[2] The first is that interpretation must attribute largely true beliefs to agents. When someone is out in the field and caught in a storm, he is more likely to believe that it is raining there rather than it is sunny. Likewise, someone is more likely to believe that human beings are mammals than they are a kind of fish. The second, which presumably forms

the core of the Principle of Charity, is that interpretation should depict the agent as conforming to rational norms, which will be the focus of detailed discussion in what follows. The third is that interpretation should represent the agent as sharing basic human values, such as a desire to survive, find love and security.

The claim that the agent should be depicted as conforming to rational norms is a development of Quine's Principle of Charity in a number of dimensions. First, Davidson's version incorporates inductive as well as deductive inference, so that the agent should be interpreted as making good inferences from evidence to hypothesis. The requirement that one should believe in a hypothesis that accords best with all the available evidence is called the Requirement of Total Evidence for Inductive Reasoning. The second respect is that Davidson's version encompasses not only theoretical reasoning, deductive and inductive included, but practical reasoning as well. Practical reasoning in its classical Aristotelian form is the inference from an end to the means that brings about that end, while in its more recent form it is the reasoning that arrives at a conclusion as to which course of action is best, all things considered. These two developments are both embraced in modern Bayesian decision theory, as explored by Ramsey and systematized by von Neumann and Morgenstern, in which theoretical reasoning is reformulated with the use of subjective probability, and practical reasoning, the inference to the action judged best all things considered, is understood as the maximization of expected utility. This essay will employ the term the *Rationality Thesis* to refer to the understanding of the Principle of Charity as requiring an agent to be sufficiently rational in these dimensions.

Closely linked to the *Rationality Thesis* is the idea that the Principle of Charity is a normative principle, which tells us not so much how an agent actually behaves as how he ought to behave. The distinction between the normative and its opposing concept, the descriptive, is a subtle one. It may at first appear that drawing the distinction implies that the Principle of Charity is not directly applicable to the actual behaviour of an agent. It is often claimed, however, that the behaviour of an agent cannot too much deviate from the rational norm, for if it does, the vocabulary of propositional attitudes will lose grip on the agent, thereby making the attribution altogether invalid. Therefore, the claim continues, even if the prescriptions of the Principle of Charity may sometimes be violated, it is nevertheless guaranteed that agents will largely conform to them.

Next to be introduced are two further theses concerning interpretation, one metaphysical and the other epistemological. They will respectively be named the *Constitution Thesis* and the *A Priori Thesis*, and are implicitly or explicitly endorsed in the writings of Davidson as well as his commentators and followers. It is in particular the latter of these that will be the main target of criticism of this essay.

The *Constitution Thesis* is a metaphysical thesis which claims that interpretation is constitutive of propositional attitudes. The concept of the metaphysical is contrasted to

that of the epistemological, and if interpretation is merely a matter of epistemology, it will be no more than a way of identifying propositional attitudes, where they are there to be found irrespective of interpretation. What the *Constitution Thesis* states is that there is nothing in propositional attitudes over and above the practice of interpretation. In this respect, interpretation should sharply be distinguished from the sciences, where the objects of scientific inquiry, such as microorganisms, genes, atoms, and black holes, although invisible to the naked eye, are accorded an existence quite independent of the activities of scientific research. They are there, whether or not any rational being sets out to discover them.

The *Constitution Thesis* is usually carefully guarded against various obvious counterarguments. It does not for example require that the possession of propositional attitudes is dependent on the agent's actually being interpreted by someone else. It suffices that the agent be interpretable, and this requirement can be made sense of without presupposing even the existence of any actual interpreter. We will not go into the details of those qualifications as they are not relevant to the following discussion.

The *A Priori Thesis*, on the other hand, is an epistemological thesis about how we know the principles governing the attribution of propositional attitudes. It affirms that the way we know them is basically *a priori*, thereby making a sharp contrast to investigations in the empirical sciences, which are obviously *a posteriori* enterprises. The claim is that, if an agent has beliefs and desires, performs actions and speaks a language, then there should necessarily be some characteristic structure that they exhibit, and that what the structure in question is is not to be determined by empirically investigating into the behaviour of agents, but by some *a priori* procedure.

The *A Priori Thesis* appears to be reached via two routs of arguments, one of which is from the *Rationality Theses*, the other from the *Constitution Thesis*. The remainder of this essay will examine each of them in turn and conclude that neither is convincing enough to persuade us of the view that the fundamental principles of interpretation can be elucidated through *a priori* philosophical analysis.

<div align="center">Ⅲ</div>

As we have seen, the *Rationality Thesis* states that the Principle of Charity sets the normative requirement that an agent should be sufficiently rational, where the kind of rationality demanded is articulated, at least partly, by Bayesian decision theory. Then the question we need to ask can be broken down into two stages. First, can the normative requirements expressed in the axioms of decision theory be formulated on *a priori* bases? And second, is it really true that decision theory, if appropriate as a normative theory, must also be sufficiently adequate as a descriptive theory? If both of these questions are answered in the affirmative, we have an argument for the *A Priori Thesis* from the *Rationality Thesis*. In the following discussion, the first question will receive an affirmative answer, while the second, which will be answered in the

negative, will be the focus of the next section.

To begin with, let us look at how the first half of the argument works with Quine's version of the Principle of Charity, which requires that an agent should not believe in contradictions in propositional and first-order predicate logic. The problem with believing in a contradiction is not limited to the practical consequences of the particular beliefs in question, but is related to the inferential rule in classical logic called the Principle of Explosion, which states that you can infer any proposition whatsoever from any pair of contradictory propositions. Now it is a truism in need of no empirical vindication that it is against any standard of rationality for an agent to believe in every proposition he can think of. The Principle of Explosion itself, too, is not something to be derived from empirical observations about how an agent forms beliefs on the basis of his other beliefs, but is part of a formal system developed by *a priori* procedures. In conjunction it follows that the rational norm that an agent should not believe in contradictory propositions is an *a priori* requirement.

For the purpose of defending the main tenets of Bayesian decision theory, two arguments, the Money Pump Argument and the Dutch Book Argument, have been offered and attracted considerable attention in the literature. A brief introduction to each of them will suffice for us to appreciate that they are both *a priori* arguments whose nature is quite similar to Quine's version of the Principle of Charity.

The Money Pump Argument is supposed to vindicate the transitivity of preference, one of the main axioms of decision theory. To say that preference is transitive means that if an agent prefers x to y, and y to z, then he prefers x to z, if we ignore for the sake of simplicity cases where one option is just as good as another. In order to see that this axiom should be accepted, imagine that an agent is facing a choice between three cars, x, y and z, where he prefers x to y, y to z, but z to x, thereby exhibiting a cyclic set of preferences. Then there is a way to exploit this set of preferences and offer choices to the agent which eventually leads to his sure loss. Suppose that he at the beginning owns z. Since he prefers y to z, there must be a sufficiently small amount of money, ε, which he will be willing to pay to have z exchanged for y. Now that he has obtained y, there must again be a sufficiently small amount of money, ε', which he will pay to have y exchanged for x, which, likewise, he will pay ε'' to be exchanged for z. Thus, at the end of all this, he is back with z, with which he started, but with a loss of $\varepsilon + \varepsilon' + \varepsilon''$ in the course of these transactions. In addition, this same procedure can be repeated an infinite number of times, through which he will be "pumped" of all his money.

A similar line of thought called the Dutch Book Argument has been advanced for the purpose of showing that subjective probability, the subjective degree of belief, should satisfy basic axioms of probability. In its simplest form, the argument runs as follows. Suppose that your subjective probability that there will be more than twenty typhoons in 2015 is 0.7. This means that you will be willing to pay up to $70 for

entering a bet a clever bookmaker offers you in which you receive $100 if there indeed are more than twenty typhoons in 2015 and nothing otherwise. Suppose at the same time that your subjective probability that there will not be more than twenty typhoons in 2015 is 0.4. Hence, you will be willing to pay up to $40 for a bet where you receive $100 if there aren't and nothing otherwise. However, if you enter both bets at the same time, you end up paying $110 for a receipt of $100, which results in the sure loss of $10. This is supposed to demonstrate that subjective probability should satisfy the axiom that the probability of *p* and the probability of not-*p* add up to 1. Similar cases can be constructed proving that all the other basic axioms of probability should equally be satisfied by subjective probability.

The parallel between these two arguments, which Petersen calls "pragmatic arguments" for the principles of decision theory, is easily visible.[(3)] In both cases it is claimed that, unless an agent conforms to certain basic rules of decision theory, he is prone to be exploited by a cleverly contrived set of choices or bets, leading to sure loss. A few observations are in order here. First, they are *a priori* arguments, in the same way as the argument for Quine's Principle of Charity is. They are not founded on empirical observations of behaviour, but rather on arm-chair theorizing about what an agent would be obliged to do under specific circumstances. Second, what they establish is the normative conclusion that an agent should respect certain principles if he does not want to lose his money for sure.

Apparently, however, violations of those principles may be quite commonplace among real-life agents, and they do not pose much serious trouble for them, because such idiosyncratic bookmakers and other exploiters are too much of a rarity in everyday circumstances. And in the very unlikely event of their arrival, an agent can then learn to adjust his judgments to conform to rational standards, in order to avoid repeated exploitation. Therefore, even if those *a priori* arguments for a normative conclusion are sound enough, their relevance to the descriptive perspective is an entirely different question, to which we will now turn.

<div align="center">IV</div>

Earlier we noted the observation that if an agent's actions deviate too much from rational norms, the vocabulary of propositional attitudes will lose its grip on him. By taking the contraposition it is equivalent to saying that as long as an agent can be described using the language of propositional attitudes, he must be sufficiently rational. Thus it seems to offer the bridge between the normative and the descriptive, which is exactly what we need in order to derive the *A Priori Thesis* from the *Rationality Thesis*.

Let us examine how promising this line of thought is by looking back at the example of the cyclic set of preferences for cars. It is possible that the agent can give an explanation for his predicament as in the following scenario. Suppose that *x* has the

most powerful engine, y comes next, and z the third, while their respective mileages are 18 km/l for x, 20 km/l for y, and 24 km/l for z. Here he comes up and says, "Well, as long as the difference in mileage is less than 5 km/h, I take that as negligible and prefer a car with a superior engine. But when the difference amounts to more than 5 km/h, I take that seriously and choose a more ecological car, irrespective of engine power."[4] This reply seems to make good sense of the fact that he is unable to make a choice between the three but yet is able to choose if the alternatives are narrowed down to two.

Suppose next that he is not able to make such a lucid explanation of what is in his mind, but instead says, "Well, I can't quite explain why, but I find it difficult to choose between the three, because if I were given a choice between x and y, I definitely prefer x, and between y and z, y, and between z and x, z. That is what I just intuitively feel." In this scenario, we may be a bit puzzled about his state of mind, but it seems possible to take his words at face value and simply note that it is a familiar characteristic of human nature that our intuitive likes and dislikes sometimes form an infinite loop of this kind that prevents us from making the final decision.

In both scenarios, it is true that the case presents some difficulties for describing and understanding the agent. First, understanding him may now depend not on his choice behaviour alone, but may require his offering some linguistic explanation, lucid or otherwise, that would make what is in his mind intelligible to us. More important, it will be impossible to represent the subjective value of each car by a number in an ordinal scale, let alone in an interval or a cardinal scale. Thus, if the numerical representation of the strengths of desires in an interval or at least in an ordinal scale is an essential component of propositional attitude attribution, then it can be argued that cyclic preference will present a serious obstacle to it and ruin our understanding of the agent.

However, if what we are after is not theoretically developing the language of propositional attitudes along decision-theoretic terms, but simply describing and understanding the agent in everyday terms and contexts, the failure to numerically represent the strengths of desires poses no threat to our goals. In daily lives we attribute beliefs and desires to agents without numerically representing their strengths, and we seem to need no guarantee that they do not exhibit cyclic preferences and other kinds of irrationalities. Thus it seems that the idea that the language of propositional attitudes presupposes the agent to be sufficiently rational appeals only to those who are already committed to a theory, a theory which we may very well suspect is derived from the normative perspective.

Here it will be illuminating to pay attention to the impressive development of behavioural economics in the last a few decades, which has cast fundamental questions on the rationality assumptions of standard economic theory. We will look at some well-known basic findings in this field, which will help us draw a sharp distinction between

the normative and the descriptive, and see that the argument for the *A Priori Thesis from the Rationality Theses* cannot go through.

Kahneman gives an interesting example in which experimental subjects systematically deviate from Bayesian reasoning.[5] He asks them about a graduate student Tom W, about whom no information is given at the initial stage, as to which department he is most likely to belong to: (1) business administration, (2) computer science, (3) engineering, (4) humanities and education, (5) law, (6) medicine, (7) library science, (8) physical and life sciences, or (9) social science and social work. The question is easy because the only key to answering it is by estimating the relative size of enrollment for each department, which will be called the "base rate". Thus humanities and education is considered more likely than computer science or library science.

Then he presents a brief personality sketch of Tom W, which runs as follows.

Tom W is of high intelligence, although lacking in true creativity. He has a need for order and clarity, and for neat and tidy systems in which every detail finds its appropriate place.... He seems to have little feel and little sympathy for other people, and does not enjoy interacting with others....[6]

Subjects are asked to rank how similar the description of Tom W is to the typical graduate student in each of the nine fields. This time, computer science, engineering, and library science are ranked higher, while humanities and education, and social science and social work are considered a poor fit. Notice that the sketch is deliberately designed in a way that makes Tom W resemble the stereotypical student of smaller departments, representing him as an anti-base-rate character.

Finally, subjects are asked to rank the fields of specialization in order of the likelihood that Tom W is a graduate student in each of them. As expected, they answer this last question exclusively on the basis of their response to the second, the similarity of Tom W's personality to the stereotype. However, this is a serious error from a statistical point of view. The base rate gives us the information for forming a prior belief about the likelihood in question, and then the Bayes's rule tells us how to update the prior belief in light of the new information about Tom W's personality. Ignoring the base rate altogether will gravely misrepresent the likelihood, and given that the Bayes's rule follows from the axioms of probability, there can be ways to exploit this error in the style of the Dutch Book Argument. However, many well-educated subjects have fallen victim to this experiment, thereby showing that the tendency to ignore the base rate is a systematic feature of human reasoning.

Another well-known finding by Kahneman concerns how agents deviate from maximizing expected utility in decision making.[7] When an agent is faced with a decision under uncertainty, expected utility theory tells him to calculate for each

alternative the weighted average of the utilities of outcomes according to their probabilities, and choose the alternative which ranks the highest. However, Kahneman has shown that people do not weigh the utility of outcomes according to their probabilities, but according to their "decision weights". Decision weight is similar to probability, as both equal to 0 when the outcome is impossible, and to 100 when it is a sure thing. But when the probability is slightly greater than 0% or slightly lower than 100%, decision weight diverges significantly from probability. Thus, the decision weight is 5.5 for 1%, while it is 91.2 for 99%. The particularly striking divergence near 100% reflects the anxiety that people feel for the possibility that an almost sure outcome may fail to obtain by an incredibly bad luck. Again, maximizing expected utility is a consequence of the formal treatment by von Neumann and Morgenstern, and therefore is the rational norm. But it does not represent how agents actually make decisions.

What is important in these observations is that in both of these examples agents deviate from rational norms in a systematic manner, and it is not as if they make sporadic mistakes here and there. On the contrary, the deviation can be studied empirically, and the study can be developed into a scientific theory about how people think and act. For example, it is a scientific finding that decision weight diverges from probability most sharply in the area close to 100%.

In addition, it seems out of question to suggest that in those cases where agents do not follow rational norms the attribution of propositional attitudes and actions to them becomes altogether invalid. Behavioural economists have no difficulty employing the language of propositional attitudes in their theories, and we have no problem taking them literally. Therefore, we here seem to have reached a firm distinction between the normative perspective and the descriptive perspective, which completes our critique of the argument for the *A Priori Thesis* from the *Rationality Thesis*.

V

The remaining task of this essay is to examine whether there is any argumentative route to the *A Priori Thesis* from the *Constitution Thesis*. Evnine, in his illuminating presentation of Davidson, says "it is *a priori* ..., and necessary, that if someone has beliefs then he generally doesn't believe open contradictions or otherwise go against the directives of these normative principles".[8] This is a typical bold statement of the *A Priori Thesis*, and the justification for this claim is the following.

> What is the relation of these principles to the mental states they govern? One picture of how they could be related is this. We identify a range of examples of actions and of states of belief, desire and so on. We then investigate these cases and find that certain principles seem to hold.... Davidson's view on how the normative principles are related to the mental is quite different. On his

view, the applicability of these normative principles is constitutively bound up
with the states they govern.[9]

He repeats roughly the same point later.

> If we take the Principle of Charity as … an empirical theory about people,
> whose beliefs and so on could be ascertained independently of it, then whatever
> criteria we took as determining beliefs and other intentional concepts, the
> Principle of Charity could turn out to be false. Davidson, however, regards
> the Principle as being constitutive of the concepts it governs. The concepts of
> belief, desire, meaning and intentional action are defined by what the 'theory',
> the Principle of Charity, says about them.[10]

In both of these places, Evnine seems to argue for the view that the Principle of
Charity is *a priori* and not empirical on the ground that it is constitutive of the
propositional attitudes that it governs. Thus here we have an argument for the *A
Priori Thesis* from the *Constitution Thesis*. He also appears to be saying that what it
means for the Principle of Charity to be constitutive of propositional attitudes is that
we cannot first identify beliefs, desires and actions independently of it and then go on
to empirically investigate their correlations in order to develop a theory about them.
On the contrary, we have the concepts to talk about propositional attitudes only when
we have the theory, the Principle of Charity, in place.

This argument, however, seems to rest on a false view of empirical theory
construction. Empirical sciences do not necessarily or even generally proceed by first
identifying various entities and then measuring the correlations between them. We can
illustrate this point by taking Mendelian genetics as an example. Mendelian theory
refers to "genes", and at least in Mendel's days they were unobservable entities.
The theory incorporates two kinds of laws, the first kind being between the genetic
constitution of parents and that of their offspring, while the second kind being between
the genetic constitution and its manifestation. Thus, it is a first kind of law that if the
father has the pair of genes AA and the mother aa, their offspring will have the pair
Aa. It is a second kind of law that having the pair of genes AA or Aa will result in
yellow leaves, while having the pair aa will result in green leaves.

Now Mendelian theory as a whole is constructed for the purpose of explaining
the observable patterns in generations of organisms, such as the pattern of yellow
and green leaves of peas. The first kind of law states the relationship between
unobservables and unobservables, while the second kind states the connection between
unobservables and observables. In neither case do we first independently identify the
entities in question and then search for correlations between them. The two kinds of
laws form a theory, and it is only when we have a theory that we can talk about and

investigate the workings of genes. Obviously, however, this procedure by no means diminishes the empirical character of Mendelian theory. On the contrary, it exemplifies an essential aspect of any scientific theory which makes reference to unobservable entities in order to explain observable patterns. Therefore, Evnine's argument for the *A Priori Thesis*, at least in the form reconstructed here, is not persuasive.

It may be complained against the preceding discussion that what is wrong here is not the argument for the *A Priori Thesis* from the *Constitution Thesis*, but rather Evnine's understanding of the *Constitution Thesis*. The point of the *Constitution Thesis* is not that we cannot talk about, identify, investigate, or otherwise think about propositional attitudes independently of the Principle of Charity, but should simply be that propositional attitudes do not exist independently of it. That, the complaint continues, is the difference between the practice of interpretation and empirical sciences. In empirical sciences, we investigate unobservable entities like genes and atoms, and the identification of those entities cannot be made independently of the respective concepts and theories. Nevertheless, genes and atoms exist there, independently of whether or not we have the requisite concepts and theories. On the other hand, what is distinctive of propositional attitudes is that they do not have the existence of their own independently of our practice of interpreting the agent. Interpretation is constitutive of propositional attitudes in this stronger sense, and if the *Constitution Thesis* is understood in this proper manner, the argument from the *Constitution Thesis* to the *A Priori Thesis* will go through.

The complaint may have a point, but it remains to be seen if it effectively reinforces the argument for the *A Priori Thesis*. The claim that in empirical sciences unobservable entities have independent existence can be taken as saying that our best available theory that incorporates concepts referring to unobservables may be false and remain to be improved, because those concepts may fail to track independently existing entities. In contrast, if propositional attitudes are constituted by interpretation, it would imply that our best present method for interpretation has no possibility of being wrong and leaves no room for improvement. But how can one make such a strong assertion about the adequacy of our way of interpreting agents? It seems that it can only be grounded on the idea that we are somehow *a priori* assured of its being the right method for understanding agents. In other words, it is only when we have some *a priori* warranty that our present practice of interpretation is on the right track that we can be sure that it leaves no room for improvement and that the *Constitution Thesis* is vindicated.

Therefore, it is rather misguided to search for an argument for the *A Priori Thesis* from the *Constitution Thesis*. Certainly, had the argument for the *A Priori Thesis* from the *Rationality Thesis* been valid, we would have had a good reason for thinking *a priori* that the Principle of Charity, with its essential emphasis on rationality, is the adequate principle for interpreting agents, and we thereby would have had a

ground for accepting the *Constitution Thesis*. But we have already seen in full detail that the argument is not convincing, given the distinction between the normative and the descriptive. In conclusion, interpretation, as a descriptive project, should be thought of as a thoroughly empirical enterprise, and there is no *a priori* way of approaching the question of what the fundamental laws governing the attribution of propositional attitudes should be.

Notes

(1) Evnine (1991), p. 104.
(2) Rescorla (2013), p. 477.
(3) Petersen (2009), p. 165.
(4) I am being unable to identify the text where Davidson, or possibly someone else, presented an example of this kind using the size of a house and its rent.
(5) Kahneman (2011), pp. 146–154.
(6) Ibid., p. 147.
(7) Ibid., pp. 314–316.
(8) Evnine (1991), p. 12.
(9) Ibid.
(10) Ibid., p. 113.

References

Evnine, S. (1991) *Donald Davidson*. Stanford University Press.
Kahneman, D. (2011) *Thinking, Fast and Slow*. Penguin Books.
Petersen, M. (2009) *An Introduction to Decision Theory*. Cambridge University Press.
Rescorla, (2013) "Rationality as a Constitutive Ideal", in E. Lepore and K. Ludwig, eds., *A Companion to Donald Davidson*. Wiley-Blackwell. 472–488.