2 Rationality and Interpretation

tukat (2007) The Thinker http://flickr.com/photos/tukat/897645495

2.1 The Varieties of Rationality

all of philosophy (...) is almost coextensive with the theory of rationality
(Putnam 1981, pp.104-105)

The detached and the engaged conception of interpretation recommend two different approaches: one based on empathic *understanding*, the other on theoretical *explanation*. Although these two conceptions frame almost all the debates regarding the nature of interpretation, another concern raises many theoretical problems: does interpretation (whether as understanding or as explanation) presuppose the rationality of the interpreted agent? And if so, to which extent? These are the issues tackled in this chapter.

Rationality is a complex, if not confused, concept. It is thus important to explore first different meanings of *rational*. First, one can distinguish **rational agents** and **rational actions/thoughts**. Rational *agents* can produce rational or irrational actions/thoughts, while non-rational or a-rational entities just cannot produce any kind of actions/thoughts. Thus rational and irrational actions/thoughts are produced by rational agents.

The actions and thoughts of rational agents are rationally assessable: they can be said to be correct or incorrect, optimal or not, justified or not, etc. Rationality is normative, and this evaluation can target either thoughts or actions. The first type is **theoretical** rationality: the rationality of inferences, thoughts, reasoning, beliefs (formation, revision, justification), mental representations, theories and explanations. The second one is **practical** rationality: the rationality of actions, behaviors, decisions, intentions, motivations, preferences and strategies. In sum, what to think *vs*. what to do. This distinction requires two sets of norms: logic, epistemology, semantics, probability theory for the first, decision theory, game theory, general equilibrium (a.k.a. market) theory for the second. Both types of rational evaluations also imply different forms of irrationality, for instance regarding arbitrary choices and wishful thinking. Regarding the former, it is theoretically irrational to believe A and not B when A and B are equiprobable, although it can be practically rational to prefer A to B when A and B are equally satisfying (not choosing can be practically irrational). Regarding the latter, it is theoretically irrational to let beliefs be influenced by desires, but it can be practically rational to let actions be influenced by desires (Harman, 1999).

Rational evaluation of practical rationality can be **internal** or **external**. In the first case, we assess the coherence of intentions, actions and plans. Actions make sense "from the point of the cognitive and conative perspective of the agent" (Stueber 2006, p. 49). In the second case, we assess the effectiveness of a rule or procedure for achieving a certain goal. Actions make sense "relative to a given set of environmental parameter that include the agent's desire but not his belief " (Bermúdez, 2002, p. 260). An action can be rational from the first perspective but not from the second one, and vice-versa. Poor performance in probabilistic reasoning can be internally rational (subjects may have good reason to choose a certain prospect) without being externally rational (their behavior is still suboptimal). The Gambler's fallacy is and will always be a fallacy: it is possible, however, that fallacious reasoners follow rational rules, maximizing an unorthodox utility function. There are thus two ways—non-exclusive—to behave irrationally. One can be externally irrational if the outcome of an action is considered suboptimal; in this case

the attribution of irrationality requires data about the agent and the outcome. (ex: Gambler's fallacy). One can also be internally irrational if, regardless of the outcomes of the action, the agent's desires and action performed are incoherent. Akrasia (acting against one's best judgment) is a form of practical irrationality because a desire (I want to stop smoking) and the action (lighting up a cigarette) are not coherent.

Many philosophers suggested that, in order to interpret another agent, a sentence or an action, we implicitly presuppose the ***rationality*** of the agent. This **assumption** modulates our interpretation and makes it possible. As Sorensen summarize the idea, "[a]nyone who superimposes the longitudes of desire and the latitudes of belief is already attributing rationality" (Sorensen, 2004p. 291). The claim can have a **weak** reading: agents are "not stupid": they have reasons to act when they behave intentionally. On a **stronger** reading, it suggest that agents should comply with rationality standards. The latter is common in classical economics: it is assumed that agents are utility-maximizer (just like physics assumes frictionless bodies or ideal gases). It can also be an **interpretative** postulate: rational agents are rationally interpretable, and rationality is a justified or useful supposition of the coherence between beliefs, desires and actions. The assumption is thought variously to be a possibility condition, an empirical hypothesis or a superfluous statement. For Davidson, it is necessary for the *application* of concepts such as action, belief, desire, intention; For Dray, it is necessary for *interpreting* action, belief, desire, intention. For Hempel (1962, p.12), it is an empirical assumptions that figures in an explanation:

A was in a situation of type C

A was a rational agent

In a situation of type C any rational agent will do x

Therefore A did x

Two important theoreticians of rational interpretation, Donald **Davidson** and Daniel **Dennett** (presented in detail in the forthcoming sections), articulated and refined the idea that rationality is also an interpretive norm (one might say a meta-norm), more fundamental than particular axioms or formal systems. For Davidson and Dennett, rationality is not a precise set of rules to follow, but a condition for being evaluated according to such rules. To be a rational agent does not imply a perfect rationality, but being assessable with rationality standards; as Stueber puts it, it is to be "normatively required to be responsive to norms of reasoning" (Stueber 2006, p. 63). In the practical domain, it supposes, among other things, that we consider the agents' beliefs and desires as reasons for action; in the theoretical domain, it supposes, among other things, that we consider the agents' beliefs and desires to be coherent.

The rationality postulate is compatible with the engaged and detached conceptions. One can argue that empathy always simulates rational agents; or that one of the assumptions of a theory of mind is the rationality of the agent. Some accounts concern the mental process of interpretation (e.g. Dennett); in this case, the rationality assumption implies that we have an intuitive concept of "rational agent" and that it is involved in interpretation. Other accounts (e.g. Dray) are about the method of the social scientist: she must presuppose the rationality of the target agent

2.2 The constitutive ideal of rationality

Davidson ideas are directed at logical positivism and Wittgensteinian philosophy. Against positivism, Davidson holds that intentional actions cannot be explained like other empirical phenomena because they involve *reasons*. Against Wittgenstein, he holds that intentional actions are also causal relations (but cannot be explained under the Deductive-Nomological model), mental objects are material objects (but cannot be described by scientific laws, a view named the "anomalism of the mental"), and that reasons are cause. Actions are particular events (such as in: the rock broke the window; no law is needed to explain it), caused by mental causes (beliefs, desires); these causes are also reasons that justify (or imply) the action when the action is rational. An irrational action is an action whose reasons causes the action but do not justify the action.

According to the **Deductive-Nomological (DN) model**, one explains an empirical phenomenon by showing how it "fit into a nomic nexus" (Hempel 1965:488). If, from the statement of a scientific law or a universal generalization (called also a covering law) and certain conditions one can logically deduce an event, this event is then regarded as being explained:

(1) L (scientific laws or universal generalization)

(2) C (condition)

---

(3) E (event)

If rationality was an empirical assumption, then action explanation would be of this kind:

If something is soluble it dissolves in liquids of a certain sort

Warm coffee is such a liquid,

All sugar is soluble,

That this cube is sugar

A particular small cube dissolved in warm coffee

This pattern of explanation requires general knowledge about sugar.

But if rationality is constitutive, then action explanations are particular explanation, grounded in local facts:

This cube was soluble

Soluble things dissolve in coffee

This cube was in coffee.

This particular small cube dissolved in coffee

Thus, to explain someone's action, we don't need (or cannot have) laws of behavior, but norms of interpretation, just like we don't need much general knowledge in the last deduction.

According to Davidson, the mental and the physical have different **constitutive** concepts: the mental in grounded in rationality. Rationality is a condition of interpretation, it is almost like a Kantian category of the understanding: it is not derived from experience, but rather makes it possible. For instance, if subjects in an experiment prefer a 100% to win 100$ than 50% to win 300$, "by the book", subjects are irrational: they prefer the option with the lowest utility. But again, certainty can have a value, therefore it is not irrational. Since it is so easy to rationalize, rationality is not an empirical concept:

> *(...) the satisfaction of conditions of consistency and rational coherence may be viewed as constitutive of the range of application of such concepts as those of belief, desire, intention and action*
>
> *(...) if we are intelligibly to attribute attitudes and beliefs, or usefully to describe motions as behaviour, then we are committed to finding, in the pattern of behaviour, belief and desire, a large degree of rationality and consistency*
>
> *(Essays on Actions and Events, p.237)*

An important point in Davidson theory, is that rationality and rational interpretation requires linguistic communication. His argument goes as the following:

> Argument:
>
> (1) To be a rational animal is to have propositional attitudes
>
> (2) Propositional attitudes are organized logically and systematically.
>
> (3) Since we consider our propositional attitudes as being true or false, correct or incorrect, we have the concepts of belief, meaning and truth
>
> (4) In order to rationalize action (as in 1), attribute propositional attitudes (as in 2) and possess the concepts mentioned in (3) a creature must be linguistic

(1) To be a rational animal is to have propositional attitudes

The argument for Davidson (1) is that when we describe the behavior of a person in terms of beliefs and desires, we show the rationality of the action in the light of the content of belief and the object of desire. The structure of the beliefs, desires and actions of Alice, in connection with her desire for apples, shows that Alice is rational because it is a coherent structure of reasons. It is therefore impossible to attribute beliefs and desires in the first place, and then, after the fact, to add that this creature is rational and to consider the last statement as informative. One cannot say that (1) "Mike took a taxi because he wanted to be in time and knew that the bus would take too much time" and saying after that (2) "Mike is rational" is informative: (2) is implicit in (1). If he has beliefs and desires, he is rational.

For example, if someone pulls a rope by its two ends, an explanation of this action that does not presuppose the rationality of the agent could be that she is fighting against herself. There are many ways to describe it as a rational action but we need to add several auxiliary hypotheses, whereas if we assume the consistency of desires, beliefs and actions, the assumption that she tries to break the rope is more natural. The attribution of propositional attitudes is thus made possible by the **constitutive ideal of rationality**

(2) Propositional attitudes are organized logically and systematically.

If one accepts (1), one has to accept (2): to have one propositional attitude is to have a panoply. The content of a belief is the sum of its relationship with all other beliefs. In order to individuate beliefs, the interpreter must locate them in a network of reasons. If an agent has the concept CAR, and that this concept specifies that a car is a vehicle, then it should be in principle possible to attribute to an agent who believes that she is near a car, the belief that she is near a vehicle. Otherwise, she just does not have the concept of CAR. If one has the concept of car, one has many general beliefs: they usually have four wheels, they go on the road, they are powered by an internal combustion engine, they are able to carry a small number of people, etc. Whatever is part of the possession of the concept CAR must figure in the set of beliefs we can attribute to this agent, otherwise it makes no sense to attribute her the concept of CAR, and we cannot make the difference between "she believes that there is a car" and "she believes that there is a big metallic object", and we cannot attribute her false beliefs, such as "she believes wrongly that this is a car while in fact it is a sculpture". IntenTional states are typically intenSional: from (1) Ana believes that Paris is in France, and (2) Paris is the capital of France, it does not follow that (3) Ana believes that Paris is the capital of France.

> *(3) Since we consider our propositional attitudes as being true or false, we have the concepts of belief and truth*

The concept of BELIEF is a necessary condition for the possession of beliefs: to believe that P implies also to believe that you believe that P. It is believed that it is believed that P when there is a possibility of surprise. If I am surprised that no coin was in my pocket, the surprise is impossible if, in the first place, I did not have a belief that there was a coin in my pocket. Thus when I am surprised, I now believe that my former belief was wrong, which supposes that I believed that I believed that I had a coin in my pocket. Surprise implies a contrast between what is considered to be true, and what is actually true, i.e. a distinction between truth and falsehood.

> *(4) In order to rationalize action (as in 1), attribute propositional attitudes (as in 2) and possess the concepts mentioned in (3) a creature must be linguistic*

Davidson argues that to be interpreted rationally, a target agent must act on the basis of propositional attitudes; however, the possession of beliefs presupposes several other beliefs, according to (3), and even the concept of BELIEF. This concept presupposes the language capability. Indeed, if one defines belief as a thought that is likely to be semantically evaluated (it can be true or false), creatures that entertain beliefs have the idea of an objective independent reality and are able to distinguish the objective and the subjective. But this distinction is guaranteed, according to Davidson, by the mastery of language. While a non-linguistic creature can at best learn and generalize, nothing in its behavior is a basis for concluding that it can distinguish the objective and the subjective. One might conclude that this is a response to different classes of stimuli, but no more. Linguistic creatures (humans) have to communicate, to be able to think that the other communicator also has the concept of an outside objective world.

Possession of a language would constitute a condition for the use of the concept of rationality, according to Davidson: "rationality is a social trait. Only communicators have it" (1982: 327). Being a linguistic agent allows us to use prior and passing theories about other speakers, and the use and revision of these theories are governed at each step by the constitutive ideal of rationality:

> *[f]or the hearer, the prior theory expresses how he is prepared in advance to interpret an utterance of the speaker, while the passing theory is how he does interpret the utterance. For the speaker, the prior theory is what he believes the interpreter's prior theory to be, while his passing theory is the theory he intends the interpreter to use. (1986, 442)*

Agents have theories about what other agents mean, and revise these theories in the course of interaction. Thus we interpret by employing our constitutive ideal of rationality. The norms of rationality are norms of interpretation and also norms that govern the use of intentional concepts.

2.3 The intentional stance

While Davidson was preoccupied by the pervasiveness of rationality in interpretation because of the open-ended possibility of justifying an action (as the practice of experimental economics show), Dennett was more preoccupied by the pervasiveness of rational interpretation displayed by cognitive science: computers, babies, animals and linguistic beings can all be described as information processors. As he notes, the notion of possession of "information or misinformation is just as Intentional a notion as that of belief" (Dennett 1971, p.90). They all can be described as wanting and desiring, but when is it correct? While Davidson was more interested by the *justification* of interpretation, Dennett is more interested by the *utility* of interpretation.

The concept of rationality, in Dennett's view, is articulated in the context of predictive strategies, intuitive attitudes or stances with respect to certain types of system with which we can interact (similar to Newell's (1990) principle of rationality). Whether it's an apple falling from a tree, a software failure or a seller of vacuum cleaners, as soon as we come into contact with these objects or agents we have a tendency to generate certain predictions about the evolution of the target object or agent (the "system"):

1. The apple, if there is no wind, will fall in a straight line toward the ground
2. The chess software will take our pawn
3. The vacuum cleaner seller will recommend that we buy one.

We have prior and passing theories for many kinds of systems. We carry out different kinds of predictions as the system is predicted and explained from the **physical stance**, the **design stance**, and the **intentional stance**. In the first case (an object which falls), our presuppositions are limited to the physical features of the objects. In the second case, our presuppositions are limited to functions (the algorithms of a computer, or the biological functions of the organs). In the last case, we presuppose the rationality of the agent, namely that the system in question is equipped with beliefs and desires:

> *"Here is how it works: first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in most instances yield a decision about what the agent ought to do; that is what you predict the agent will do." (Daniel Dennett, 1987, p. 17)*

> *"The success of the stance is of course a matter settled pragmatically, without reference to whether the object really has beliefs, intentions, and so forth; so whether or not any computer can be conscious, or have thoughts or desires, some computers undeniably are intentional systems, for they are systems whose behavior can be predicted, and most efficiently predicted, by adopting the intentional stance toward them." (Dennett 1978, p. 238)*

The strategy can be employed with many different systems: computers, babies, animals, humans, etc. What matters is the *success* of the stance. Dennett gives the example of the plover and its "deceptive" behavior when it detects a predator; it behaves as if its wing was broken, and "pretends" to be injured. Is the deception a real intentional deception, which reveals the rationality of the agent, or a simple reflex? To examine this (Dennett, 1987: 258) proposes to interpret the behavior of the plover as a soliloquy, where the animal intentionally tries to fool the predators. If we can successfully use the intentional stance, if it shows that the animal is sensitive to a complex set of environmental conditions that the soliloquy suggests, we can say that it entertains beliefs and desires. If these tests show that cognitive control mechanisms react to simple environmental patterns, then it is necessary to reassess its rationality downwards. Computers are another interesting case: we say that the computer *wants* to take my pawn, but this ascription is useful or not useful, not right or wrong: it is just more convenient to say that than to describe all its algorithms and internal states. This procedure is the same, according to Dennett, as we use between ourselves: we assume that agents are informed and capable of logical inferences, predict their behavior, and reassess their rationality if necessary.

Stances are not appropriate and relevant for any kind of system:

the physical stance is not useful when we want to talk about the functioning of a software

the design stance is not useful when we talk about the fall of an object

the intentional stance is not useful when we talk about the movement of clouds, or the trajectory of a missile

It is useful when it picks up regular patterns. On the other hand, we can adopt "the Intentional stance in one's role as opponent, the design stance in one's role as redesigner, and the physical stance in one's role as repairman" (1971, p. 91).

Thus the intentional stance is both a cognitive mechanism and a social-scientific methodology. It is grounded in natural selection: if evolution has done its job, our prediction will be useful (1971, p.93). Our folk-psychology is thus an adaptation. Social science must proceed by assuming adapted design, and going from the intentional to the design stance. Instead of restricting intentionality to language, Dennett restricts it to (useful attribution of) information-processing. One can call it *evolutionary rationalism*: natural selection designed us to "get the answer right and to want to get it right" (97). Thus beliefs, desires and so on are like centers of gravity: not "Real", but useful, like gravity centers in physics.

2.4 Ideals and Stances

Both Dennett and Davidson are continuers of Quine and the pragmatist tradition, but with different inspirations and aspirations. Davidson faced the problems of the irreducibility of rationalization in experimental economics; Dennett saw how the intentional stance works in so

many circumstances with humans and non-humans. Davidson focuses on interaction with other humans, Dennett with other "systems"; Davidson on justification, Dennett on usefulness. They have different conceptual grounds: for Davidson it is logic and semantics, for Dennett, evolutionary biology and cognitive science (where non-linguistic agents are intentional agents). They also have different explanatory "target": linguistic communicator vs. information-processing systems. They also had different impact: Davidson on social science and literary studies, Dennett on cognitive science and biology.

However, both take rationality as a fundamental, *sui generis* notion (linked with beliefs, desires, prediction and explanation) that cannot be completely codified: it is an ideal notion, but this ideal is not the economist's rational agent. It is more like the warrant of commonsense psychology. It is "a good sense of when to rely on what" (Dennett, 1987, p.97). Beliefs and desires are "abstracta". There are no basic principles of rationality (Davidson, 2004 :196). Both see beliefs and desire ascriptions as interlocked, and attributing rationality as a faster way of interpreting. For Davidson, it is faster than irrational interpretation (e.g. see the rope example). It is easier to see it as a rational action than a non-rational one. For Dennett, it is faster than physical stance interpretation (e.g. predicting a computer's move based on quantum physics). Both see irrationality as meaningful only in the light of rationality. For Davidson, one has to suppose some partition in the mind ("semi-independent structures" see (Davidson, 1982b)). For Dennett, we cannot attribute irrationality from the intentional stance, but need the intentional stance to understand malfunctioning psychological mechanisms and their role in our holistic cognitive economy. Both see questions of interpretation as ubiquitous and as an everyday competence based on an intuitive theory.

One of their main differences is about mental and non-mental property. Davidson opposes the mental and the physical, as two domains based on different constitutive concepts. Dennett see that there are 3, not 2, domains (the physical, the functional and the intentional), and that although they have their own properties, one can use a higher-level domain to investigate the lower-level. Davidson sees intentionality, rationality and normativity as tied together. No language, no mind. Dennett advocates a continuist position. They also have different stances on psychological explanation: for Dennett we can go on and mechanically decompose intentional systems in subsystems. For Davidson the holism, the anomalism of the mental and its irreducibility to the physical makes that impossible. This difference is probably because Dennett links intentionality (and rationality) to information rather than language and semantics. Both see the problems of over-rationalization (attributing rationality to non-rational system such as missiles or flowers). For Davidson, it is unjustified, for Dennett, it is not useful.

Both are more on the **detached** side. For Davidson, interpreters adjust their prior/passing theories and the mental can only be understood through particularist explanation (where no generalizations are involved), as in "the window broke because it was struck by a rock" (Davidson, 1963). We rationalize actions with beliefs and desires under the ideal of rationality, but not causes. For Dennett, when we interpret, we have a theory of behavior (1971, 93):

> *"An interesting idea lurking in Stich's view is that when we interpret others we do so not so much by theorizing about them as by using ourselves as analog computers that produce a result. Wanting to know more about your frame of mind, I somehow put myself in it, or as close to being in it as I can muster, and see what I thereupon think (want, do. . .). There is much that is puzzling about such an idea. How can it work without being a kind of theorizing in the end? For the state I put myself in is not belief but make-believe*

> *belief. If I make believe I am a suspension bridge and wonder what I will do when the wind blows, what "comes to me" in my make-believe state depends on how sophisticated my knowledge is of the physics and engineering of suspension bridges. Why should my making believe I have your beliefs be any different?* "(1987, P. 100)

2.5 References

Baker, L. R. (1989). Instrumental Intentionality. Philosophy of Science, 56(2), 303-316.

Davidson, D. (1963). Actions, Reasons, and Causes. The Journal of Philosophy, 60(23), 685-700.

Davidson, D. (1973). Radical Interpretation. dialectica, 27(3-4), 313-328.

Davidson, D. (1982). Rational Animals. dialectica, 36(4), 317-327.

Davidson, D. (1982b). Paradoxes of Irrationality. In R. Wollheim & J. Hopkins (Eds.), Philosophical Essays on Freud (pp. 289-305).

Davidson, D. (1986). A Nice Derangement of Epitaphs. In Ernest Lepore (Ed.), Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson (pp. 433-446). Oxford: Basil Blackwell.

Davidson, D. (2004). Problems of Rationality. Oxford: Oxford University Press.

Dennett, D. C. (1971). Intentional Systems. The Journal of Philosophy, 68(4), 87-106.

Dennett, D. C. (1978). Brainstorms : Philosophical Essays on Mind and Psychology (1st ed.)Montgomery, Vt.: Bradford Books.

Dennett, D. C. (1987). The Intentional Stance. Cambridge, Mass.: MIT Press.

Dennett, D. C. (1990). The Interpretation of Texts, People and Other Artifacts. Philosophy and Phenomenological Research, 50, 177-194.

Dray, W.H. "The Historical Explanation of Actions Reconsidered," p. 106 in William H. Dray (ed.), Philosophical Analysis and History. New York: Harper & Row, 1966.

Follesdal, D. (1982). The Status of Rationality Assumptions in Interpretation and in the Explanation of Action. Dialectica, 36(4), 301-316.

Hempel, C. G. ([1961]2001). Rational Action. In J. H. Fetzer (Ed.), The Philosophy of Carl G. Hempel: Studies in Science, Explanation, and Rationality (pp. 311-328). Oxford: Oxford University Press.

Hempel, C. G. (1965). Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science. New York,: Free Press.

Harman, G. (1999). Reasoning, Meaning, and Mind. Oxford: Oxford University Press.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, Mass.: Harvard University Press.

Popper, K. R. (1994). Models, Instruments, and Truth: The Status of the Rationality Principle in the Social Sciences. In The Myth of the Framework. In Defence of Science and Rationality (pp. 154-184). London: Routledge.

Putnam, H. (1981). Reason, Truth, and History. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

Sorensen, R. (2004). Charity Implies Meta-Charity. Philosophy and Phenomenological Research, 26, 290-315.