Rationality

Michael Rescorla

Abstract: A striking thesis lies at the core of Davidson's philosophy: when we attribute intentional content to another creature's mental states and speech acts, we must treat the creature as largely conforming to our own rational norms. I will discuss how this thesis informs Davidson's treatment of rationality and intentionality. After reviewing some historical background, I present basic aspects of Davidson's position. I then examine various worries about the position. I conclude by highlighting some key Davidsonian insights into rationality.

A striking thesis lies at the core of Davidson's philosophy: when we attribute intentional content to another creature's mental states and speech acts, we must treat the creature as largely conforming to our own rational norms. I will discuss how this thesis informs Davidson's treatment of rationality and intentionality. After reviewing some historical background (§1), I present basic aspects of Davidson's position (§2). I then examine various worries about the position (§§3-6). I conclude by highlighting some key Davidsonian insights into rationality (§7).

§1. Kant, Carnap, and Quine

How does logic relate to thought? Three mutually consistent answers surface repeatedly in the philosophical tradition:

- Logic is *normative*, in that logical laws dictate how one should think.

- Logical laws *describe* how people actually think, at least to a first approximation.

- Logic is *constitutive* of thought, meaning roughly that conformity to logical laws fundamentally informs what it is to think.

Note that the third answer entails the second.

Kant regards logic as both normative and constitutive. He begins the *Jäsche Logic* by emphasizing normativity (1800/1974, 16): "In logic, we do not want to know how the understanding is and thinks, and how it hitherto has proceeded in thinking, but how it ought to proceed in thinking." But Kant also assigns logic a crucial descriptive role within his *a priori* theory of human mental faculties. He postulates faculties of *understanding* and *sensibility*. The understanding cannot deviate on its own from logical norms: "if we had no other power of cognition besides the understanding, we would never err" (1800/1974, 59). Logical errors result from "the *unnoticed influence of sensibility upon the understanding*" (1800/1974, 59). Logic describes how the understanding operates when not illicitly disrupted by sensibility. Logic delineates "necessary laws of the understanding and reason in general," without which "we could not think at all" (1800/1974, 14-15). In that sense, logic is constitutive of thought.

In the early 20th century, philosophers such as Wittgenstein and the logical positivists rejected Kant's approach as overly *psychologistic*. Yet many of those same philosophers advanced *depsychologized* versions of the Kantian thesis that logic plays a constitutive role within rational inquiry.

To illustrate, consider Carnap's *The Logical Syntax of Language*. Carnap wants to replace traditional philosophy with *rational reconstruction*. He urges us to render scientific discourse rigorous and precise by constructing *linguistic frameworks*. In delineating a linguistic framework, we specify the logic governing inquiry within that framework. There is no meaningful question regarding which frameworks are correct or incorrect: "*In logic, there are no morals*. Everyone is at liberty to build up his own logic, i.e. his own language, as he wishes" (1937/2002, 52). We can offer *pragmatic exhortations* regarding the merits of one framework over another. We can say that one framework is simpler, more elegant, and so on. But we cannot rationally assess frameworks, because our framework provides our only applicable standards of rational assessment. Carnap eschews all talk about mental faculties, mental processes, and the like. In that respect, he differs fundamentally from Kant. Nevertheless, Carnap pursues a *linguified*, *relativized* version of the Kantian thesis that logic plays a constitutive role within rational inquiry: *linguified*, because logic now helps constitute a linguistic framework; *relativized*, because there are diverse equally legitimate frameworks.

In the 1950s, Quine launched a battery of arguments against Carnap's approach. Carnap sharply distinguishes *change within a framework* (as when one revises an estimate of some object's mass) from *change between frameworks* (as when one revises a logical law). According to Quine, Carnap has not provided any clear explanation of this difference. He has not isolated a clear sense in which certain doctrines play a constitutive role within one's current scientific theorizing. Quine sees no principled difference between revisions in logic and other revisions in our scientific theory (1953/1980, 43). If we change our logic, we are not altering constitutive aspects of thought or discourse. We are merely altering particularly well-entrenched elements of our overall science. Thus, Carnap's picture does not describe even idealized scientific inquiry.

Quine concludes that Carnapian rational reconstruction lacks philosophical interest: "why all this creative reconstruction, all this make-believe? The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology?" (1969, 75). Quine advocates *naturalized epistemology*: an abstract branch of scientific psychology. Naturalized epistemology studies how mental activity transforms sensory input into theories of the world. Rather than reconstruct human rationality within an artificial linguistic framework, we study actual human mental processes. "Better to discover how science is in fact developed and learned than to fabricate a fictitious structure to a similar effect" (1969, 78).

Quine's approach seems to jettison normativity from serious discourse. Naturalized epistemology describes *actual* mental activity. What role can normative prescriptions occupy within this enterprise? As Kim (1988, 389) complains, "Quine is urging us to replace a normative theory of cognition with a descriptive science." If we seek only to describe actual mental activity, then how can we say that logic sets norms for *correct* thinking? More generally, how can we normatively evaluate beliefs as *justified* or *unjustified*?

Another pressing worry concerns the relation between naturalized epistemology and the mind's representational capacity. Quine finds no place for representationality (or *intentionality*) within his version of naturalized epistemology. In *Word and Object* (1960, 26-79), he advances his *radical translation* thought experiment. He imagines a linguist attempting to translate an unknown language by observing the linguistic behavior of native speakers. Quine contends that there are multiple incompatible translation manuals equally consistent with total observed linguistic behavior. Furthermore, he contends that there is *no fact of the matter* regarding which of these translation manuals is correct. On this basis, Quine urges us to jettison intentionality

from scientific discourse (1960, 216-221), relying instead upon non-intentional explanations drawn from physics, biology, neuroscience, or Skinnerian behaviorist psychology.

Quine's view threatens to banish normativity and intentionality from our theorizing. This threat provides the immediate context for Davidson's work (Friedman 1996). Davidson accepts many aspects of Quine's approach, but he wants to secure a more robust role for normativity and intentionality. In particular, he seeks to preserve the traditional conception of humans as rational.

§2. Davidsonian interpretation theory

Davidson emphasizes *interpretation*: our practice of ascribing intentional content to one another's linguistic performances and mental states. Interpretation is central to daily human interaction. For example, we frequently explain someone's actions by attributing beliefs and desires to her. Davidson supplements Quine's austere naturalist picture with a systematic theory of intentional ascription.

Following Quine, Davidson (1984, 125-154) imagines a linguist studying an unfamiliar language. Davidson modifies Quine's thought experiment in two key respects:

- Whereas Quine considers radical *translation*, Davidson considers radical *interpretation*. The goal is to *interpret* sentences, rather than to translate them into one's own language (1984, 128-129). More specifically, Davidson's radical interpreter seeks to construct a Tarski-style truth-theory for the native language (1984, 130-131).
- Whereas Quine emphasizes *language*, Davidson places equal emphasis upon *mind*. As Davidson notes (1984, 142-145), it is relatively straightforward to determine what someone believes and desires if we already know what her words mean, or to

determine what her words mean if we already know what she believes and desires. The radical interpreter must solve for all three variables simultaneously: beliefs, desires, and meanings.

The centerpiece of Davidson's philosophy is an idealized model of radical interpretation.

Davidson grounds his model in *Bayesian decision theory*, as developed by Ramsey (1926/1950) and refined by Jeffrey (1983). Bayesian decision theory is a mathematical model of decision-making under uncertainty. Bayesians codify belief through *subjective probability* and desire through *utility*. Probabilities and utilities determine *expected utilities* for actions available to the agent. A Bayesian agent chooses actions that maximize expected utility. To render this mathematical apparatus more concrete, Bayesians typically prove a *representation theorem*: if an agent's preferences satisfy certain constraints (such as transitivity), then there exist probabilities and utilities with respect to which her preferences maximize expected utility. In Ramsey's treatment, preferences determine unique probabilities, and they determine utilities uniquely up to linear transformation. In Jeffrey's treatment, preferences determine utilities only up to fractional linear transformation, and they determine probabilities only to within a certain quantization. Bayesian representation theorems show how to extract a semi-unique theory of mental states (probabilities and utilities) from a relatively observable evidentiary base (preferences). The theorems thereby illuminate how one can read a rational pattern into observed behavior.

Traditional Bayesian decision theory presupposes meaning-theoretic facts. For example, Jeffrey assumes that the agent has preferences over *propositions*. Davidson complains that meaning-theoretic presuppositions are not available to a radical interpreter (2004, 29, 160). Accordingly, he generalizes Jeffrey's model. In Davidson's later work, the radical interpreter takes as data *whether the subject prefers one sentence true over another* (2004, 161). More

precisely, the radical interpreter observes the distal conditions under which the native prefers one sentence true over another.[1] From that data, the radical interpreter must construct a unified theory that specifies probabilities, utilities, and a Tarski-style truth-theory. In this way, radical interpretation grounds belief, desire, and meaning in observed behavior.

To develop his model, Davidson deploys Quine's *Principle of Charity*: "assertions startlingly false on the face of them are likely to turn on hidden differences of language" (1960, 59). Citing Charity, Quine urges that translation should not render speakers as denying basic logical truths (1960, 58-59). Davidson concurs, and he vastly generalizes the point. According to Davidson, radical interpretation should construe natives as conforming to basic logic, probability theory, and decision theory. Assuming that the agent conforms to basic axioms of probability and decision theory, the radical interpreter can identify which native locutions correspond to familiar truth-functional connectives (1990, 326-328). Once we have identified truth-functional connectives, we can apply the Bayesian representation theorem to discern the agent's probabilities and utilities (1990, 328). We then apply Charity once more: we assume that the agent conforms to basic quantificational logic, and we thereby identify quantifiers, predicates, and singular terms as such (1990, 319-320). The standard Tarskian apparatus determines how to interpret logical vocabulary. Thus, our only remaining task is to interpret primitive singular terms and predicates. Davidson's remarks regarding this crucial task are not as systematic as one might desire (1984, 136-137, 151-152), (1990, 320-322), (1999, 82), (2001, 148-149), (2004, 126). The basic idea is to apply a form of Charity. However, Davidson employs widely varying formulations to express the requisite kind of Charity.[2] A recurring theme is that, if the native assigns high probability to a sentence, then we should take this as *prima facie* evidence that the

sentence is true. Another recurring theme is that we should strive for an interpretation on which the native's conditional probabilities track genuine relations of evidential support.

Obviously, ordinary people do not execute anything resembling Davidsonian radical interpretation. But Davidson maintains that his model enshrines a *constitutive ideal of rationality* underlying all interpretation, radical or otherwise (1980, 222-223). The model makes explicit certain normative constraints that necessarily govern intentional ascription (2004, 128). We can attribute intentionality to a creature only if we treat the creature as largely conforming to rational norms. *Which* rational norms? Our own, because those are the only ones we have. We can describe a creature's intentional states *as* intentional only if we postulate that the creature largely satisfies our own standards of rationality (2005, 319):

> Charity is a matter of finding enough rationality in those we would understand to make sense of what they say and do, for unless we succeed in this, we cannot identify the contents of their words and thoughts. Seeing rationality in others is a matter of recognizing our own norms in their speech and behavior. These norms include norms of logical consistency, of action in reasonable accord with essential or basic interests, and the acceptance of views that are sensible in the light of evidence.

An agent need not always satisfy rational norms. However, any lapses are deviations from an overall rational pattern (2004, 196):

> [I]t does not make sense to ask, concerning a creature with propositional attitudes, whether that creature is *in general* rational, whether its attitudes and intentional actions are in accord with the basic standards of rationality. Rationality, in this primitive sense, is a condition of having thoughts at all. The question whether a creature "subscribes" to the principle of continence, or to the logic of the sentential calculus, or to the principle of

total evidence for inductive reasoning, is not an empirical question… An agent cannot

fail to comport most of the time with basic norms of rationality.

Thus, intentional psychology is an inherently normative enterprise (1980, 241):

[T]here is no way psychology can avoid consideration of the nature of rationality, of

coherence and consistency… Psychology, if it deals with propositional attitudes… cannot

be divorced from such questions as what constitutes a good argument, a valid inference, a

rational plan, or a good reason for acting.

Intentional description and normative evaluation are inextricably entangled.

In this manner, Davidson revives the Kantian thesis that logical norms are constitutive of

thought. As Davidson puts it, "I think of logic and decision theory as rough but essential laws of

thinking and action," in that they "delineate aspects of rationality which thinking creatures must

to a considerable extent exemplify" (1999b, 620). Davidson's distinctive twist on the Kantian

thesis is that logical norms, along with other rational norms, are constitutive of *interpretation*.

For Kant, logical norms flow from Aristotelian logic. For Davidson, they flow from classical

first-order logic with identity. Whereas Carnap countenances diverse legitimate linguistic

frameworks, Davidson allows no such diversity: "all thinking creatures subscribe to *my* basic

standards or norms of rationality" (Davidson 2004, 195). Kant, Carnap, and Davidson all diverge

from Quine by assigning a special constitutive status to logic.[3]

The contrast between Quine and Davidson emerges in their differing treatments of the

logical connectives. Quine holds that we should project our logic onto the natives, but he stresses

that being "[b]eing thus built into translation is not an exclusive trait of logic. If the natives are

not prepared to assent to a certain sentence in the rain, then equally we have reason not to

translate the sentence as 'It is raining'" (1970/1986, 82). Quine's guiding maxim is that we

should not render natives as denying obvious truths, whether those truths are logical,

observational, or otherwise. By Gödel's completeness theorem, each logical truth is either

obvious or derivable from obvious truths through individually obvious steps. Yet many non-

logical truths are also obvious, such as "1+1=2," or "It is raining" uttered while it is raining.

Quine denies that logic has any privileged status vis-à-vis other obvious truths.[4]

Davidson does not ground charity in the "obviousness" of logic. Instead, he assigns logic

a central role in *constituting* mental content (2004, 156-7):

> We individuate and identify beliefs, as we do desires, intentions, and meanings in a great
>
> number of ways. But relations between beliefs play a decisive constitutive role; we
>
> cannot accept great or obvious deviations from rationality without threatening the
>
> intelligibility of our attributions. If we are going to understand the speech or actions of
>
> another person, we must suppose that their beliefs are incorporated in a pattern that is in
>
> essential respects like the pattern of our own beliefs. First, then, we have no choice but to
>
> project our own logic on to the beliefs of another.[5]

It is not entirely clear how to interpret this passage, or other similar passages in Davidson's

writings (2004, 97-98, 138-139, 196). Evidently, though, Davidson wants to secure a distinctive

constitutive status for logic. He holds that we can intelligibly impute propositional content to

mental states only if we treat those states as largely conforming to our own logical norms. One

might say that Davidson tries to sanitize constitutivity by subtly emending Quinean charity.[6]

§3. Charity as a constraint on interpretation

Davidson classifies several distinct interpretative constraints under the label "Principle of

Charity," including the following:

(1)   Interpretation must ascribe a background of *true* beliefs to the speaker: "it is impossible for the interpreter to understand a speaker and at the same time discover the speaker to be largely wrong about the world" (2001, 150).

(2)   Interpretation must depict the speaker as largely conforming to the interpreter's own rational norms, including norms given by logic, probability theory, and decision theory.

(3)   Interpretation must depict the speaker as largely sharing the interpreter's own *values*. For example, we must assume that the speaker "shares with us a desire to find warmth, love, security and success" (2004, 183).

Commentators have criticized all three constraints. I focus upon (1) and (2).

A common objection to (1) is that it seems possible for a subject to be systematically mistaken about her surroundings (Burge 2003, 336-337), (Goldman 1986, 175-176). As Lewis (1983, 112-113) observes, good interpretation must allow that experience can mislead. To take an extreme example, suppose we are interpreting a recently envatted brain trapped in a Matrix-style computer simulation. Quite plausibly, the best interpretation will construe her as massively deluded about her external environment. The brain's experiences are systematically misleading, so her beliefs about the external world are systematically mistaken.[7]

To motivate (1), Davidson writes: "sentences that express… beliefs, and the beliefs themselves, are correctly understood to be about the public things and events that cause them, and so they must be mainly veridical" (2001, 174). Similar passages recur frequently in his writings. Burge (2003) suggests that such passages overemphasize the role played by *the subject's own current causal connections to the world* in fixing mental content. Quite plausibly, representational mental activity presupposes a baseline of accurate representation. Quite plausibly, representation of the distal environment arises only through representationally

successful causal transactions with the environment. But those transactions might have occurred earlier in the subject's history, as with a recently envatted brain. Burge (2010, 69) suggests that the transactions might even have occurred much earlier in the subject's evolutionary history. There is no evident reason why good interpretation must regard *the subject herself* as largely correct in her *current* beliefs.[8]

Philosophers have also proposed various counter-examples to (2). Bortolloti (2005) suggests that delusional subjects routinely flout basic rational norms. Setting aside such extreme cases, it is unclear whether normal humans conform to rational norms as closely as Davidson suggests. Research by Kahneman and Tversky (1974) and many other psychologists indicates that normal human reasoning and decision-making display numerous deficiencies (Rysiew, 2008): the gambler's fallacy, framing effects, the anchoring effect, and so on. Citing this research, Goldman (1989), Nozick (1993, 152-156), and Thagard and Nisbett (1983) argue that Davidson overstates the centrality of rational norms to intentional ascription. Some philosophers even question whether basic norms of consistency are as central to interpretation as Davidson intimates. Citing the paradox of the preface, Goldman (1989) denies that we should strive to interpret subjects as consistent.

How damning are these putative counter-examples to (2)? Davidson acknowledges that one can violate rational norms. Indeed, he carefully analyzes phenomena such as akrasia, inconsistent belief, and self-deception (1980, 21-42), (2004, 167-230). He would surely urge that the putative counter-examples are *deviations* from background conformity to rational norms (Ludwig 2004, 348-349). Unfortunately, Davidson never articulates his own position very precisely. He never states how closely a thinking creature must conform to rational norms.

§4. Realism, instrumentalism, and eliminativism

Quine acknowledges that intentional locutions play a central role in daily interaction. He denies that they deserve any place in scientific discourse. He insists that we should eschew intentional talk when "limning the true and ultimate structure of reality" (1960, 221). In that sense, Quine is an *eliminativist*. Churchland (1981) and Stich (1983) share Quine's eliminativism. Burge (2010) and Fodor (1987) endorse an opposing *intentional realism*: intentional locutions denote genuine properties that we should cite within scientific psychology. Dennett (1987) occupies an intermediate instrumentalist position. He agrees that "the intentional stance" plays a useful predictive role in scientific theorizing, but he questions whether mental states *really* have intentional properties.

How does Davidson's position relate to these debates? Davidson professes to be an intentional realist (2001, 70-84). He writes that "propositional attitudes… are every bit as real as atoms and baseball bats, and the facts about them are as real as the facts about anything else" (2005, 316). Nevertheless, readers often interpret Davidson as an intentional anti-realist, albeit one exquisitely attuned to the central role that intentional attribution occupies within our lives. Williamson (2004, 137) discerns in Davidson's work an "ideal verificationism, on which agents have just the intentional states that a good interpreter with unlimited access to non-intentional data would ascribe to them." Burge (2003, 359-360) attributes to Davidson "the view that an interpreter's interpretation helps constitute the linguistic and mental content of the creature being interpreted." These readings suggest a less than fully realist posture towards intentionality and rationality. Intentional properties are tied to intentional ascription procedures, in a way that physical properties are not tied to procedures for ascribing physical properties. So intentional facts are somehow less "real" than non-intentional facts.

Several facets of Davidson's position suggest anti-realist sympathies. Most notably, Davidson follows Quine in holding that radical interpretation is *indeterminate*: the same linguistic data admit several equally good interpretations, between which "there may be no objective grounds for choice" (1980, 222). As Davidson emphasizes, Jeffrey's decision theory mandates limited indeterminacy regarding the assignment of probabilities and utilities to sentences (1990, 323-324, fn. 66). Davidson also foresees indeterminacy surrounding the interpretation of sentences. For example, he claims that one might build equally good interpretations around either of the two interpretation clauses (2001, 78-79):

"Rome" denotes Rome

"Rome" denotes an area 100 miles to the south of Rome

He even suggests that equally good interpretations may differ regarding the truth-values of certain utterances, such as borderline color ascriptions (2001, 80-81). Davidson does not simply claim that the data *underdetermine* intentional ascription (2001, 75-76). Underdetermination of theory by evidence afflicts all scientific theorizing, not merely intentional psychology. Davidson furthermore claims that there is "no fact of the matter" regarding which interpretation is correct (2001, 214). In that sense, intentional ascription is genuinely *indeterminate*.[9]

Davidson insists that indeterminacy of interpretation is perfectly consistent with intentional realism. He repeatedly adduces an analogy with measurement (2001, 75). We can measure weight using either pounds or kilograms. No one would deny on that basis that weight measurements describe real features of the world. Similarly, he says, the indeterminacy of interpretation should not lead us to deny the reality of intentional facts.

It seems to me that the measurement analogy undercuts rather than supports Davidson's professions of intentional realism. There is no conflict between saying that an object weighs 10

pounds and saying that it weighs 4.5 kilograms. These are simply two ways of describing the same fact. In contrast, there is a *prima facie* conflict between saying that "Rome" denotes Rome versus an area 100 miles south of Rome. There is a *prima facie* conflict between theories that assign different truth-values to the same utterance. Davidson claims that such *prima facie* conflicts are sometimes illusory. He claims that diverging interpretations are sometimes compatible (2001, 76). But it is difficult to see how genuine intentional realists can agree. (Cf. Lepore and Ludwig 2005, 239-247, 382-386; Rawling 2001, 244-250.)

Another anti-realist tendency in Davidson's writings is his steadfast focus upon intentional *attribution*. Davidson emphasizes idealized procedures for *discovering* mental content. His core methodology is to "adopt the stance of a radical interpreter when asking about the nature of belief" (2001, 148). From an intentional realist perspective, this methodology is puzzling. According to intentional realism, mental states have their contents independently of any theorist's procedures for discovering those contents. Intentional phenomena no more depend upon our theorizing about intentionality than physical phenomena depend upon our theorizing about the physical. Fodor and Lepore (2007, 687) pointedly express this realist viewpoint:

> Does anybody *still* think that… a theory of (radical) translation/interpretation… would illuminate questions in semantics or in the metaphysics of meaning? Interpretation and translation, as Davidson and Quine understood them, are to be construed in *epistemological* terms… What has the epistemology of interpretation got to do with the metaphysics of content?

Davidson's persistent tendency to move between the metaphysics and the epistemology of intentionality strongly suggest a kind of intentional anti-realism.[10]

§5. Cognitive science and radical interpretation

Let us now consider more carefully the scientific status of psychology. Davidson holds that intentional psychology differs from physics and biology in a crucial respect: intentional attribution is governed by the constitutive ideal of rationality. While cautioning that "it would be meretricious to summarize these points by saying that psychology… is not a science," he insists that "psychology is set off from other sciences in an important and interesting way" (1980, 241). He also urges that "[t]he propositional attitudes do not seem suited to incorporation into a unified scientific view of the world" (2001, 71). After posing the question "Could there be a science of rationality?" (2004, 117), he answers that he "does not know, nor much care" whether "a psychological theory is so different from a theory in the natural sciences as not to deserve to be called a science" (2004, 134). Overall, these passages suggest a broadly Quinean picture: we delineate a *first grade* theory, comprising physics, biology, and other "hard" sciences; and we also delineate a *second grade* theory that includes intentional psychology (Quine 1969, 24). Davidson explores the second grade theory in much more detail than Quine. Ultimately, though, Davidson seems to share Quine's wariness towards the scientific credentials of intentionality.[11]

A good illustration is Davidson's dismissive attitude towards cognitive science. A few exceptions aside, Davidson almost entirely ignores contemporary scientific research into intentional mental activity.[12] Rather than discuss the science of intentionality, Davidson emphasizes his own idealized armchair model of radical interpretation.

Why this make-believe? Why consider how an idealized radical interpreter might attribute intentional content to mental states? Why not simply study intentional mental activity from a scientific perspective? Few contemporary philosophers would suggest that we should study the physical world by reconstructing how one might come to know physical facts. Our best

strategy for illuminating the physical is to analyze physics as it currently operates. Likewise, isn't our best strategy for illuminating the mental to analyze our current best science of mental activity? Why study how an idealized theorist might discover intentional facts, rather than studying what current science tells us about intentional states and processes?

Cognitive science offers well-confirmed theories that explain various mental phenomena. Contrary to Quinean strictures, those theories assign intentionality a central role (Burge 2010), (Fodor 1989). For example, perceptual psychology studies how the perceptual system transits from proximal sensory input to a percept that represents the distal environment as being a certain way (Burge 2010), (Knill and Richards 1996). Similarly, empirical linguistics offers impressive theories of the semantic competence deployed during linguistic comprehension (Heim and Kratzer 1998); some of those theories even borrow Davidson's emphasis upon Tarski-style semantics (Larson and Segal 1995). Cognitive science routinely individuates mental states in representational terms. It thereby yields numerous insights into intentional mental activity. Why ignore those insights?

To develop this challenge, Fodor and Lepore (1994) emphasize a particularly important point: Davidsonian radical interpretation employs a severely impoverished evidentiary base. Radical interpretation takes as data *only* the distal circumstances under which the speaker prefers one sentence true over another. No cognitive scientist would accept any such draconian evidentiary restriction. Depending on the explanatory context, cognitive scientists cite numerous additional evidentiary sources: intentional descriptions of speech acts; syntactic or semantic features of other human languages; discoveries about the mental activity of non-human species; neural facts; and so on. What interest attaches to a philosophical model that deliberately ignores so many potentially valuable sources of evidence?

Davidson responds that his project is orthogonal to scientific inquiry. His project is to elucidate "what it is about propositional thought --- our beliefs, desires, intentions, and speech --- that makes them intelligible to others" (2004, 133). How is it possible to extract beliefs, desires, and meanings from observed linguistic behavior? To answer this question, Davidson reconstructs interpretive practice. His reconstruction highlights how "the normative character of thought, desire, speech, and action imposes [structure] on correct attributions of attitudes to others, and hence on interpretations of their speech and explanations of their actions" (1990, 325).

In theory, Davidson's response sounds reasonable enough. Embracing an ecumenical spirit, one might pursue armchair analysis of radical interpretation *in addition to* scientifically informed philosophizing about the mind. In practice, however, Davidson's prioritization of armchair reconstruction over empirical science arguably exerts a distorting influence upon his philosophy. I provide an example in the next section.

§6. Representation and rationality in non-linguistic creatures

Radical interpretation applies only to creatures that speak a language. This restriction to linguistic creatures reflects broader Davidsonian commitments. Davidson repeatedly urges that non-linguistic animals lack anything resembling the cognitive capacities displayed by linguistic creatures: "to be a thinking, rational creature, the creature must be able to express many thoughts, and above all, be able to interpret the speech and thoughts of others" (2001, 100). Intentional content arises through a network of cognitive capacities available only to language-speakers (2001, 100):

> a very complex pattern of behavior must be observed to justify the attribution of a single
>
> thought. Or, more accurately, there has to be good reason to believe there is such a

complex pattern of behavior. And unless there is actually such a complex pattern of

behavior, there is no thought. I think there is such a pattern only if the agent has

language.

Thus, "rationality is a social trait. Only communicators have it" (2001, p. 105).

Davidson offers several arguments for this assessment (1984, 162-164, 167-170), (2001,

97-98, 102-105, 117-121, 128-134), (2004, 135-149). The arguments have gained few adherents.

Burge (2010, 264-283) and Ludwig (2004) urge that the arguments are problematic. In my

opinion, Davidson has provided no compelling reason to suspect that thought requires language.

Davidson claims that thought must have logical form: "language and thought require the

structure provided by a logic of quantification" (2004, 140). He also claims that logically

structured mental states arise only when a creature can execute rudimentary logical inference.

Even if we concede these two claims, there is no obvious reason to believe that thought requires

language. There is no obvious reason why linguistic capacities are necessary for executing

logical inferences over logically structured representational mental states. Despite millennia of

discussion, no philosopher has yet provided a convincing argument that logical capacities

presuppose mastery of a natural language.

Also problematic is Davidson's exclusive focus upon logically structured mental

representation. Even if we grant that *thought* requires logically structure, there is no obvious

reason why *mental representation* requires logical structure. Citing perceptual psychology,

Burge (2010, 537-540) argues that *perceptual states* have representational content but not logical

structure. Similarly, many psychologists propose that animals navigate by exploiting *cognitive*

maps: mental representations that operate roughly like concrete maps (Gallistel 1990), (Tolman

1948). On one plausible view, cognitive maps have compositionally significant *geometric*

structure but not *logical* structure (Rescorla 2009a, b). Davidson says nothing to address, let alone rebut, the possibility of non-logical mental representation.

Nor is there any obvious reason why *rationality* requires logical capacities. Bayesian decision theory does not presuppose logically structured representations. Admittedly, Ramsey and Jeffrey formulate Bayesian decision theory over logically structured entities. But one can just as easily develop Bayesian decision theory over non-logical representations. For example, Bayesian inference can operate over cognitive maps that lack logical structure (Rescorla 2009a, b). A hypothetical creature could navigate by executing Bayesian updating over these maps. Such a creature would conform to *Bayesian* norms but not *logical* norms. Actual non-human animals may well navigate in this way. Davidson provides no compelling argument that rational norms of probability and decision theory presuppose linguistic competence.

A basic problem facing Davidson is that cognitive science routinely attributes representational mental activity to non-linguistic creatures, including not only pre-linguistic infants and but also non-human animals. Many non-linguistic creatures perceptually represent distal properties such as shapes, sizes, and colors (Burge 2010, 342-366, 419-430). Many non-linguistic animals navigate by representing the surrounding spatial environment (Burge 2010, 492-518), (Cheng, Shettleworth, Huttenlocher, and Rieser 2007). Non-human primates display advanced cognitive capacities, including capacities to represent social dominance relations (Cheney and Seyfarth 2007) and to execute impressive spatial reasoning (Call 2006). The scientific literature amply supports these and numerous other representational attributions. Yet Davidson's position seemingly forbids representational attributions to non-linguistic creatures. Davidson does not acknowledge, let alone try to resolve, the apparent conflict between his position and contemporary science.

§7. Davidson's contributions to the study of rationality

A venerable philosophical tradition seeks to conscript logic and probability theory into the service of psychological description. This tradition waned in the mid-20th century, partly due to prevailing behaviorist tendencies, partly because philosophers took so much to heart Frege's advice that we "always separate sharply the psychological from the logical" (1884/1980, x). For example, Carnap systematically ignores actual human mental processes, instead focusing his energy upon rationally reconstructing deductive and inductive reasoning (1950, 37-51, 576).[13] Although Quine rejects Carnapian rational reconstruction, he inherits Carnap's distaste for rational modeling of mental processes. Quine, even more than Carnap, evinces strong behaviorist hostility to the mental.

Davidson helped revive the traditional emphasis upon rational modeling of mental activity. He sparked renewed philosophical research into the descriptive psychological import of rational norms by placing the following proposal at center stage:

Logic, probability theory, and decision theory set norms for intentional mental activity.

Those norms should guide systematic psychological description.

Davidson develops the proposal primarily through armchair reconstruction of interpretive practice, rather than through detailed study of empirical psychology. But we can detach the proposal from Davidson's favored armchair methodology.

Particularly noteworthy is Davidson's prescient focus upon Bayesian decision theory as a descriptive paradigm. Over the past few decades, cognitive science has embraced the Bayesian paradigm. Perceptual psychology is an especially impressive illustration (Knill and Richards 1996). According to Bayesian perceptual psychology, the perceptual system estimates distal properties through a Bayesian inference prompted by proximal sensory input (Rescorla

forthcoming). As this research program demonstrates, numerous perceptual illusions result from mental activity that conforms (at least approximately) to Bayesian norms. Cognitive scientists have successfully extended the Bayesian paradigm beyond perception to diverse phenomena, including *sensorimotor control* (Bays and Wolpert 2007), *language parsing* and *acquisition* (Chater and Manning 2006), "central" cognitive processes such as *concept acquisition* and *causal reasoning* (Chater and Oaksford 2008), and non-human *navigation* (Cheng, Shettleworth, Huttenlocher, and Rieser 2007). In each case, the Bayesian research program aims to establish that relevant mental activity conforms (at least approximately) to Bayesian norms.

Typical work within this research program features two main elements: a normative model of a mental task; and empirical confirmation that the normative model approximately describes actual mental activity. For example, perceptual psychologists delineate how an ideal Bayesian agent would estimate an object's shape based upon proximal sensory input. Empirical investigation reveals that the resulting normative model approximately describes actual human shape perception. A similar template underlies Bayesian modeling of human cognition. In this spirit, Oaksford and Chater (2009) examine various apparent defects in human logical reasoning, and they argue that the apparent defects actually reflect the operation of near-optimal Bayesian inductive inference. The Bayesian research program embodies a broadly Davidsonian methodology: it seeks to match empirical mental phenomena as well as possible to normative models. Indeed, Oaksford and Chater (2009) cite Davidson as an antecedent.

I submit that Davidson isolates a scientifically fruitful explanatory strategy, even though he himself largely ignores cognitive science. The strategy remains controversial. Many cognitive scientists deny that psychological explanation should consult normative models (Elqayam and Evans 2011). Nevertheless, much of the best current psychological research embodies a broadly

Davidsonian entanglement of normative evaluation and psychological description. If this research is on the right track, then intentionality and normativity should occupy a central role within any decent naturalized epistemology.

Davidson's work also advances several intriguing metaphysical doctrines that bear upon the putative entanglement of normative and descriptive factors:

- Intentional mental states arise only through baseline conformity to elementary rational norms.

- Logically structured intentional mental states arise only through baseline conformity to elementary logical norms.

Both doctrines require careful formulation, so as to avoid possible counter-examples like those discussed in §3. But it seems plausible that doctrines along the foregoing lines are correct. Philosophers should continue to investigate these and similar doctrines. By doing so, we might elucidate the descriptive success of normative models within empirical psychology. We might also clarify what it is for mental states to have intentional content.

References

Bays, Paul, and Daniel Wolpert. 2007. "Computational Principles of Sensorimotor Control that Minimize Uncertainty and Variability." *Journal of Physiology*, 578: 387-396.

Bortolloti, Lisa. 2005. "Delusions and the Background of Rationality." *Mind and Language*, 20: 189-208.

Burge, Tyler. 2003. "Reply to Stroud." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, edited by Martin Hahn and Bjørn Ramberg. Cambridge: MIT Press.

Burge, Tyler. 2010. *Origins of Objectivity*. Oxford: Oxford University Press.

Call, Josep. 2006. "Descartes's Two Errors: Reason and Reflection in the Great Apes." In *Rational Animals?*, edited by Susan Hurley and Matthew Nudds. Oxford: Oxford University Press.

Carnap, Rudolf. 1937/2002. *The Logical Syntax of Language*. Translated by Amethe Smeaton. Chicago: Open Court Press.

Carnap, Rudolf 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Carnap, Rudolf. 1971. "Inductive Logic and Rational Decisions." In *Studies in Inductive Logic and Probability*, vol. 1, edited by Rudolf Carnap and Richard Jeffrey. Berkeley: University of California Press.

Chater, Nick, and Christopher Manning. 2006. "Probabilistic Models of Language Processing and Acquisition." *Trends in Cognitive Science*, 10: 335-344.

Chater, Nick, and Mike Oaksford, eds. 2008. *The Probabilistic Mind*. Oxford: Oxford University Press.

Cheney, Dorothy, and Robert Seyfarth. 2007. *Baboon Metaphysics*. Chicago: University of Chicago Press.

Cheng, Ken, Sara Shettleworth, Janellen Huttenlocher, and John Rieser. 2007. "Bayesian Integration of Spatial Information." *Psychological Bulletin*, 133: 625-637.

Churchland, Paul. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy*, 78: 67-90.

Davidson, Donald. 1974. "Replies to Lewis and Quine." *Synthese*, 27: 345-349.

Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Davidson, Donald. 1990. "The Structure and Content of Truth." *Journal of Philosophy*, 87: 279-328.

Davidson, Donald. 1999a. "Reply to Quine." In *The Philosophy of Donald Davidson*, edited by Lewis Hahn. Chicago: Open Court.

Davidson, Donald. 1999b. "Reply to Bjørn Ramberg." In *The Philosophy of Donald Davidson*, edited by Lewis Hahn. Chicago: Open Court.

Davidson, Donald. 2001. *Subjective, Intersubjective, Objective*. Oxford: Clarendon Press.

Davidson, Donald. 2004. *Problems of Rationality*. Oxford: Clarendon Press.

Davidson, Donald. 2005. *Truth, Language, and History*. Oxford: Clarendon Press.

Dennett, Daniel. 1987. *The Intentional Stance*. Cambridge: MIT Press.

Elqayam, Shira, and Jonathan Evans. 2011. "Subtracting 'Ought' From 'Is': Descriptivism Versus Normativism in the Study of Human Thinking." *Behavioral and Brain Sciences*, 34: 233-290.

Fodor, Jerry. 1987. *Psychosemantics*. Cambridge: MIT Press.

Fodor, Jerry, and Ernie Lepore. 1994. "Is Radical Interpretation Possible?". *Philosophical Perspectives*, 8: 101-119.

Fodor, Jerry, and Ernie Lepore. 2007. "Brandom Beleaguered." *Philosophy and Phenomenological Research*, 74: 677-691.

Frege, Gottlob. 1884/1980. *The Foundations of Arithmetic*, 2nd. revised ed., trans. J. L. Austin. Oxford: Basil Blackwell.

Frege, Gottlob. 1893/1967. *The Basic Laws of Arithmetic: Exposition of the System*, trans. M. Furth. Berkeley: University of California Press.

Friedman, Michael. 1996. "Exorcising the Philosophical Tradition: Comments on John McDowell's *Mind and World*." *Philosophical Review*, 105: 427-467.

Gallistel, C. R. 1990. *The Organization of Learning*. Cambridge: MIT Press.

Goldman, Alvin. 1986. *Epistemology and Cognition*. Cambridge: Harvard University Press.

Goldman, Alvin. 1989. "Interpretation Psychologized." *Mind and Language*, 4: 161-185.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in Generative Grammar*. Malden: Blackwell.

Jeffrey, Richard. 1983. *The Logic of Decision*. 2nd. ed. Chicago: University of Chicago Press.

Kant, Immanuel. 1800/1974. *Logic*. Trans. R. Hartman and W. Schwarz. New York: Dover.

Kim, Jaegwon. 1988. "What is Naturalized Epistemology?". *Philosophical Perspectives* 2, edited by J. Tomberlin. Atascadero: Ridgeview.

Knill, David, and Whitman Richards, eds. 1996. *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Larson, Richard, and Gabriel Segal. 1995. *Knowledge of Meaning*. Cambridge: MIT Press.

Lepore, Ernie, and Kirk Ludwig. 2005. *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: Clarendon Press.

Lewis, David. 1983. *Philosophical Papers*, vol. 1. Oxford: Oxford University Press.

Ludwig, Kirk. 2004. "Rationality, Language, and Charity." In *The Oxford Handbook of Rationality*, edited by A. Mele and P. Rowling. Oxford: Oxford University Press.

Nozick, Robert. 1993. *The Nature of Rationality*. Princeton: Princeton University Press.

Oaksford, Mike, and Nick Chater. 2009. "Précis of *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*." *Behavioral and Brain Sciences*, 32: 69-120.

Quine, W. V. 1953/1980. *From a Logical Point of View*. 2nd. ed., revised. Cambridge: Harvard University Press.

Quine, W. V. 1960. *Word and Object*. Cambridge: MIT Press.

Quine, W. V. 1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Quine, W. V. 1970/1986. *Philosophy of Logic*. 2nd. ed. Cambridge: Harvard University Press.

Pearson, James. 2011. "Distinguishing W. V. Quine and Donald Davidson." *Journal for the History of Analytical Philosophy*, 1: 1-22.

Ramsey, Frank. 1926/1950. "Truth and Probability." Rpt. in *Foundations of Mathematics*. New York: Humanities Press.

Rawling, Piers. 2001. "Davidson's Measurement Theoretic Reduction of the Mind." In *Interpreting Davidson*, edited by Petr Kotatko, Peter Pagin, and Gabriel Segal. Stanford: CSLI Publications.

Rescorla, Michael. 2009a. "Chrysippus's Dog as a Case Study in Non-Linguistic Cognition." In *The Philosophy of Animal Minds*, edited by Robert Lurz. Cambridge: Cambridge University Press.

Rescorla, Michael. 2009b. "Cognitive Maps and the Language of Thought." *British Journal for the Philosophy of Science*, 60: 377-407.

Rescorla, Michael. Forthcoming. "Bayesian Perceptual Psychology." In *The Oxford Handbook of the Philosophy of Perception*, edited by M. Matthen.

Rysiew, Patrick. 2008. "Rationality Disputes --- Psychology and Epistemology." *Philosophy Compass*, 3: 1153-1176.

Stich, Stephen. 1983. *From Folk Psychology to Cognitive Science*. Cambridge: MIT Press.

Thagard, Paul, and Richard Nisbett. 1983. "Rationality and Charity." *Philosophy of Science*, 50: 250-267.

Tolman, Edward. 1948. "Cognitive Maps in Rats and Men." *Psychological Review*, 55: 189-208.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185: 1124-1131.

Williamson, Timothy. 2004. "Philosophical 'Intuitions' and Skepticism about Judgment." *Dialectica*, 58: 109-153.

Biographical Note: Michael Rescorla is an Associate Professor in the philosophy department at

the University of California, Santa Barbara. His research concerns the philosophies of mind,

language, and logic.

Notes

[1] By emphasizing distal conditions, Davidson departs substantially from Quine, who instead emphasizes patterns of *proximal* sensory stimulation. As Davidson stresses, the contrast has significant implications for epistemology and for the study of mental content (1999, 82-84), (2001, 151), (2005, 47-62).

[2] See (Lepore and Ludwig 2005, 182-197, 258-259) for discussion.

[3] Frege imagines discovering creatures "whose laws of thought flatly contradicted ours and therefore frequently led to contrary results even in practice" (1893/1967, 14). Frege holds that such creatures would display "a hitherto unknown type of madness," but he seems to think that such creatures are possible in principle. In contrast, Davidson would deny that Frege has denied a possible scenario. He would say that the creatures are not interpretable and that they do not have intentional mental states.

[4] Quine officially defines "obvious" in idiosyncratic behaviorist fashion: by calling a statement "obvious to a community," he explains, "I mean only that everyone, nearly enough, will assent to it, for whatever reason" (1970, 82). This is not what "obvious" means in any ordinary usage. It is unclear why translation should try to preserve "obviousness" *in Quine's official sense*. I think that Quine's rationale for charity trades upon the normal epistemological connotations of "obvious," despite Quine's official behaviorist definition.

[5] What is it to "project" our logic onto the natives? In the ensuing text, Davidson glosses "projection" as follows: we assume that the native's beliefs are "logically consistent (up to a point at least)" (2004, 157). Other passages suggest that interpretation requires a stronger kind of logical projection: namely, we assume a native tendency to reason in accord with certain basic deductive rules (2004, 97, 138-139, 195-196).

[6] Quine sometimes appears to allow that an alien language could deviate radically from first-order quantificational logic, thereby rendering the language untranslatable by us. Davidson denies that any such language is possible (1999a, 81-82). For discussion, see (Pearson 2011).

[7] Davidson (1974, 346) concedes to Lewis that good interpretation should allow for "explicable error." I think that Davidson does not pursue this concession to its natural consequence: that "explicable errors" can be so massive and systematic as to undercut anything resembling (1). As Lepore and Ludwig (2005, 329-332) emphasize, Davidson presupposes something like (1) when deploying charity against skepticism (2001, 150-153).

[8] Burge (2010, 68) advances a de-epistemologized, weakened variant of (1) that takes these points into account.

[9] Davidson defends the indeterminacy thesis through several distinct arguments. One argument, exemplified by the "Rome" example, maintains that one can alter the satisfaction relation while holding truth-conditions constant (1984, 227-241). Another argument claims that there are several alternative ways of solving simultaneously for beliefs and meanings. Radical interpretation "must separate meaning from opinion partly on normative grounds" by applying norms of rationality (2001, p. 215); "various norms can suggest conflicting ways of interpreting an agent," and "there may be no clear grounds for preferring one of these ways to others" (2005, 319).

[10] Lewis (1983) employs radical interpretation as a literary device for dramatizing a *metaphysical* problem: reducing the intentional to the non-intentional. In Davidson's hands, the *epistemological* aspects of radical interpretation seem far more crucial.

[11] In his first papers defending anomalous monism, Davidson (1980, 222-223) draws an invidious distinction between physics and intentional psychology. He argues that the former but not the latter can include strict (i.e. exceptionless) laws. He defends this conclusion by citing the constitutive ideal of rationality. Fodor (1987, 5-6) retorts that geology, biology, and other special sciences include non-strict (i.e. *ceteris paribus*) rather than strict laws. Davidson eventually concedes the point (2005, 191-193). Still, he insists that intentional psychology is distinguished from other sciences through its distinctive entanglement with normative evaluation (2004, 114).

[12] In the 1950s, Davidson conducted experimental research designed to test decision theory empirically. He eventually concluded that decision theory was not subject to empirical test: intentional content is attributable to a subject's mental states only if we presuppose that the subject largely conforms to our own rational norms, including

decision-theoretic norms (1980, 236-239). On this basis, Davidson critiques some putative empirical counter-examples to decision theory (1980, 270-273).

[13] In posthumous work, Carnap countenances using probability theory as a tool for psychological description (1971, 11-12). Even here, he devotes no serious attention to the descriptive import of normative modeling. In contrast, Davidson's philosophical system is an extended meditation on the relation between psychological description and normative evaluation.