The mind we do not change[1]

Wolfram Hinzen

To appear in Olsson, E. (ed.), The philosophy of Isaac Levi, forthcoming in *Cambridge Studies in Probability, Induction and Decision Theory*, Cambridge University Press (series editor: Brian Skyrms).

1. Introduction

For many years Isaac Levi has been a staunch defender of a strictly normative and prescriptive conception of rationality. The origin and motivation for this crucial commitment, as it transpires particularly clearly in *The Covenant of Reason* (Levi 1997; henceforth CR), has been Levi's exploration and development of the Peirce-Dewey "belief-doubt" model of inquiry. On the latter, justifiable change in state of belief is a species of rational decision-making. This is what motivates Levi's concern with "rationality" in the first place (CR, 20). In fact, no substantive commitment on what rationality substantively *is* -- or on what it is to be rational -- emerges from this theoretical interest. In particular, we are not told what beliefs or values we should have, which ones it is rational to have, or how we should base our beliefs on "evidence". Rather, principles of rationality are primarily justified instrumentally through their *regulative use* as formal constraints on well-conducted inquiry and problem-solving, no matter the domain, be it science, politics, economics, technology, or art, or even simply the personal decisions we face in daily life. Given their exceeding generality, we can only expect constraints on the coherence of choice to be both formal and weak. Principles of rationality are to be kept immune from revision if a general theory of how rational changes in point of view are to be justified is to be possible at all (CR, 24). But I understand this is to be an essentially practical necessity, which does not

---

[1] Dedicated to Isaac Levi, who changed my mind completely during memorable years in New York. But it changed again -- in ways this essay documents. We should be looking for something stable!

depend on a notion of what the "essence" of rationality is. We are dealing with a fundamentally *instrumental* conception of rationality here (CR, 16), not with a conception in which rationality is something to strive for or to analyse for its own sake.

I find much to admire in this vision, whose at times quite radical minimalism and modesty as regards the study of rationality contrasts quite sharply with more portentous conceptions of it (and of us as "essentially rational beings"): e.g., it seems to offer little support for the idea that the theory of rationality can be appealed to in an effort to explain and "rationalize" the political and economic organization of modern societies, say as the forming of a form of "contract" between naturally constituted rational individuals confronting each other as competitors for scarce resources in a state of nature. On a different score, despite its decidedly narrow focus Levi's vision of rationality has clear and ramified implications for the agenda of 20th century philosophy, not only with regard to metaphysical issues of correspondence and reference or the nature of propositions (cf. Levi 1991), but also with regard to the issue of meaning, the analytic, and the apriori. The best parts of the latter, one might argue Levi's viewpoint to imply, fall out from an account of how our revisions of belief are constrained (so that analytic truths, in particular, would be an epiphenomenon of the fact that beliefs have varying degrees of entrenchment).

All that said, I will use this opportunity to take a step back and read Levi somewhat against himself, confronting his vision of philosophy and rationality with another, more naturalistic one, in ways that may not only illuminate it, but also change it internally. Particularly if rationality is fundamentally instrumental, naturalizing rationality seems an option, contrary to what Levi suggests.[2] There is, I emphasize, no question that metaphysical issues such as

---

[2] Probably *only* *if* this is so. I see little scope for a naturalization of intrinsic (rather than instrumental) value, and Dennett's (1995, Chs. 16-17) "naturalized ethics" as a failure in this respect.

naturalization are peripheral to Levi's main concerns. Even epistemological issues have an unclear status, if these, say, include debates over the correctness of empiricist versus pragmatist or rationalist so-called "theories of knowledge". Levi, while of course a committed pragmatist, does not actually give us a "theory of knowledge", especially if this includes a conceptual definition of what knowledge is (cf. Levi 1980, henceforth EK, section 1.9). An analysis of the "Enterprise of Knowledge" -- a theory of justified *change* of belief or states of knowledge -- is a quite different enterprise. Still, I will argue that discussing features of both naturalism and rationalism helps bringing important features of Levi's philosophy clearer into view. The bottom-line is that while it is true, of course, in one sense, that we "change our minds" (how we should do so being Levi's lifelong theme), there is also the mind we do not change: the (rational) mind we happen to have, by virtue of our evolution and nature.

2. Levi's Normativism

On Levi's view, the theory of rationality is not in the service of telling us something about the natural world. Rather than being a descriptive theory concerned with what is *true*, it sets a *standard* that may guide the deliberating agent in the monitoring of his own decisions and changes of belief, while having neither explanatory nor predictive uses. Since there is no limit to human confusion, we will typically not live up to this standard. In this very sense, being rational cannot be a property that empirically characterizes us as humans. As is more generally recognized in discussions of bounded rationality in Bayesianism, poor memory, limited computational capacities within given time constraints, acting on conflicting motives without resolving the conflicts inherent in them, emotional stress, lack of self-knowledge, issues of identity and religious value not reducible to material consumptions, etc., will necessarily lead us to act "irrationally" in the light of received views of rational choice in classical economic and Bayesian economic theory. The disconfirmation of rational choice theory understood as an

empirical one has been argued to be one of the truly robust results achieved in the human sciences (Conlisk 1996).

Thus it cannot be rationality that makes us agents. Levi's solution to this problem for classical rational choice theory is that we are not the "rational being", but the being that *tries* to be rational, setting itself and using a standard of rationality for evaluating its own actions (cf. CR, 6-7).

Levi's normativism ties in naturally and clearly with an explicit verdict against naturalism, in that something is rational only if you *evaluate* it according to a certain standard, and there is no road leading from facts to values (CR, 14). The *prescriptive* does not reduce to the *physical*. No natural being *can* as such be rational. Since Levi effectively and interestingly equates the mental or intensional with the prescriptive -- though he makes the point almost in passing (CR, 14) -- the mind never comes into view as a natural object with its own structure and function. It rather only *arises* from the evaluation of certain natural facts, such as utterances and actions. In evaluating them, they come to be seen as either *generating* certain commitments -- for example, by saying what he does the agent becomes committed to be disposed to endorse certain consequences -- or as *fulfilling* them -- thus I may evaluate the handing-over of an amount of money as the fulfilment of a promise or contract. Drawing up agreements for a contract, in Levi's terms, (necessarily) involves events that are describable in purely physical terms, such as the putting-down of a pen on paper.[3] But using a standard of rationality, we can re-describe these very events as generating an obligation, a re-description in which prescriptive moral principles and laws of contract are necessarily involved.

---

[3] The example is Levi's, and I myself would want to note already here that terms like "pen" and "paper" have no status in physical theory, hence are not either describable in physical terms alone.

Note that it is not *quite* the case, here, that there is nothing to say about the mind in an empirical and naturalistic (or "psychological") perspective. On the contrary, human rational behaviour is subject to empirical study. Crucially, however, we can study empirically only what we evaluate normatively too. There can be no such question as whether or how rational an agent behaved, without a view on what it would have been for that agent to have behaved rationally in this circumstance -- that is, without viewing his actual actions as a *performance* realizing certain attitudinal *commitments*. With the view on the agent's commitments in place, his degree of fulfilment of the latter through his actions becomes an empirical question. Commitment and performance are correlative on this view, showing that there is no empirical dimension to the study of human action and mentality without a normative one. As Levi puts it in a somewhat Kantian mood (CR, 16), commitment without performance is "empty" -- the rational agent must be viewable as attempting to realize his commitments, even if he does not live up to them -- but performance without commitment is "blind" -- for understanding of an agent's actions comes when we can see his actions as attempts to fulfil his commitments.[4]

Levi's strict normativism is premised on the unbounded nature of rationality, that is, its non-relativity to the limits of our capacity to fulfil rational obligations we incur. It is the unbounded nature of rationality that makes us rational failures, and deprives the theory of rationality of an empirical (explanatory and predictive) content.[5] A more provocative way of putting this same insight is that the empirical study of the human mind reduces to the *clinical* study of its rational deficiency. No theoretical insight into the structure of rationality needs to

---

[4] Although Levi's reference to Kant is only in passing, there is, I believe, a much more systematic connection here between Levi's non- or anti-naturalism and a Kantian "two-worlds"-view. Much of his emphasis on conflicts between a first-person and a third-person perspective on choice, or between the autonomy of the agent and the predictability of its trajectory viewed as a naturalistic entity (see section 6), relates to a Kantian mode of thought (cf. also Rabinowicz 2002, 91-2).

[5] "Principles [of rational belief, desire and choice] fail to contribute to explanation of behaviour as physically described because such principles are false as applied to human beings." (CR, 7, fn. 2; cf. 36-7)

flow from this. A *practical* need arises, rather, to improve, through therapy, training, and prosthetic devices, the agent's performance against the background of his commitments. Much in Levi's philosophical frame revolves around the distinction between *inquiry* -- called for when a change of belief is to be justified -- and *therapy* -- called for when an agent does not realize which dispositions he is committed to have (CR, 11).

It is the verdict against bounded rationality, however, that I would take least issue with here. I find it plausible that if we clearly see what a rational solution to a decision problem is, and still do not decide in practice in accordance with this insight, there is no reason here to lower our standards. As long as a therapy can be sought that brings us closer to a rational ideal, I do not see what should lead us to abandon the ideal *as such*. Suppose indeed we lower the standards. Then

"no matter how we trim our principles of rationality, there will always be predicaments so complex and stressful as to preclude the applicability of the eviscerated standard. Evisceration will continue until nothing of interest is left to carve out." (CR 8)

While having little to add to this argument (which makes bounded rationality approaches self-refuting), it seems to me that there are others. Arguably, modified versions of classical rationality designed to cope with anomalies blurring the predictive power of classical rationality tend to retain a basic orientation to the classical conception of full rationality. The theorist of bounded rationality will typically test which assumptions lead to behaviour different from what unbounded rationality predicts. In this sense, as Foley (2003) remarks,

"bounded rationality is an epicyclic extension of rationality theory, and struggles to define itself in terms other than its deviations from the results of full substantive rationality" (p. 4)

Given the heterogeneity of causes for anomalies it seems unclear how a unified theory of bounded rationality can be achieved, "which instead tends to degenerate from an explanatory framework into a descriptive language" (ibid.).

Levi's framework, by embarking on strict normativism, avoids these pitfalls. This becomes clear already in the decision to model beliefs as *commitments*, for commitments are by their nature not inconsistent with a failure to live up to them (a broken promise is still a promise, after all). Rather than concluding that the classical theory of rationality is simply a disconfirmed theory, it lifts its laws to the level of regulatory principles for self-criticism that have a normative status only. Rational beliefs are structured according to the very strong logic involving the condition $B(A) \rightarrow A$, that whatever the agent fully believes is true actually is. Doxastic coherence demands this, Levi argues (CR, 66). The one thing that saves us from absurdity here is to give up on viewing that logic as a logic characterizing a set of truths (hence as describing the real world). The logic of this kind of coherence, in terms adapted from Ramsey, is the "logic of consistency", not the "logic of truth" (CR, 44-5).

But while there is nothing to object to the beauty and ingeniousness of this solution, it also leaves us in the dark about what an explanatory framework for the explanation of human action should be. Figuring out which therapeutic treatments would do best for a given agent struggling with rationality is not a task catching everyone's ambition. And should we content ourselves that there is little scope for a naturalistic inquiry into the rational structure of the human mind at all? That, indeed, there is no specifically "human" mind to speak of, given that, on Levi's view, a normative re-description of physical events according to standards of rationality can be provided for the actions of more abstract agents such as institutions and governments as well?

3. Naturalism and rationalism

Levi's picture of the mind could hardly be in greater conflict with the rationalist tradition, as I have myself construed it in my own ways in a forthcoming book (Hinzen, in press, a). On this alternative view, man's reason is an expression of his nature, and to the extent that man's rational mind can bring light into the structure of the real, reality itself is rationally structured. Human reason is subject to empirical and naturalistic study -- our minds are structured by grammatical, moral, aesthetic, etc. knowledge -- and in pursuing this study there is no intrinsic need to hold human performance against a prescriptive norm.[6] The human mind and its intrinsic structure is as such what is rational in nature, and its rationality has no more to do with how we evaluate nature than other aspects of the natural world. Stipulated norms on human behaviour and public policies deriving from contingent values have to be held against human nature as a constraint on which norms and policies we should have. The science of morality is, as on the Humean (1739-40) conception, part of the "science of human nature", a study of an aspect of human beings qua natural beings, in Hume's conception an inherent part of the new sciences of his day. Moral psychology becomes the empirical study of our species-specific faculty of reasoning and moral judgement, cartographing its structure and content.[7]

While being outmoded, the scientific basis of this picture in our own time seems clear enough.[8] Cognitive science has identified and described structures inherent in the human mind that enable our various cognitive competences to develop in the uniform way they do (Mehler and Dupoux 1994, Hirschfeld and Gelman 1994, Pinker 2003, Bjorklund and Pellegrini 2002).

---

[6] Quite the contrary, I tend to agree with Wertheimer's (and Goodman's) "Factunorm" principle, according to which how we do think is necessarily evidence for the principles of rationality. "We are (implicitly) accepting the Factunorm Principle whenever we try to determine what or how we ought to think. For we must, in that very attempt, think." (Wertheimer 1974, 110-1).

[7] Harman (1998) suggests just such an agenda for moral theory.

[8] The scientific basis for the Quinean project of a "naturalization of philosophy", for which Skinner's radical empiricism was instrumental, remains unclear in comparison. From the viewpoint of traditional rationalist or Chomskyan rationalism, Quinean "naturalism" is the opposite of that (Hinzen, 2004, in press a).

Naturally both the methodologies and the results of these works are open to revision and empirical refutation. What matters is they require empirical counterarguments, whereas claims that human nature does not exist, which abound in the philosophical literature (e.g., Rorty 1990, evoking and echoing a long tradition of continental philosophy including Heidegger, Gadamer, and Derrida), are usually not argued for on empirical grounds, and are often simply taken for granted.[9]

Human nature, for all empirical analyses suggest, exists in the sense of intrinsic and species-specific structures characterizing the human mind. While the human language faculty is a classical example, more recent studies suggest that infants have an innate expectation regarding the nature of objects (objects move as bounded wholes, they are solid, they move on connected paths, they may be inanimate or animate, etc.). A recent study suggests that 5-month-old infants do not readily apply physical principles to humans that they apply to physical and inanimate objects, and in fact have trouble viewing them as material objects at all. Overall, the study concludes,

"young infants may have different modes of construal for humans versus inanimate objects: humans are construed in terms of social and intentional actions, while inanimate objects are interpreted via a system sensitive to object physics" (Kuhlmeier, Bloom and Wynn, in press).

---

[9] In the analytic tradition, equally, it seems that no notion of human nature has ever centrally figured. Frege and Wittgenstein drew philosophers' attention away from natural language as an intrinsic property of the human mind. For Carnap, the only exception to the basic correctness of (a pragmatist version of) empiricism was our knowledge of logic. Quine thought the attempt to study, on an empirical basis, the innate structures of the mind that enter into human language use a form of "folly". It seems no exaggeration that the mind as the rationalist -- in Quine's case, it was Chomsky's reincarnation of it -- proposed to study it played virtually no role in the philosophical reflection on the mind ever since Frege, Russell, Wittgenstein, Carnap and Quine set 20th century philosophy on its course.

Construing actions as social includes, according to some views, construing them as normative (or as subject to contractual obligations, as in Cosmides and Tooby's classical 1992 analysis). To the extent that these empirical conclusions are sensible, one explicit motive for Levi's anti-naturalism disappears. As noted, Levi takes Brentano's thesis of the irreducibility of the intensional to the physical on board, collapsing it with the fact-value dichotomy. Somehow, the idea seems to be, naturalistic science can deal with what it is to put a pen down on paper, but it cannot deal with what it is to take up an obligation (as by signing a contract). But from the viewpoint of human beings (infants, indeed), the world is not first or primarily "physical", and then the intentional comes somehow in addition to it by these humans adopting certain values for re-describing nature. For all developmental psychology suggests, the intentional does not have to do with contingent values at all. It has to do with what creatures we are, and what internal structures we bring, unconsciously, to bear on the reality we happen to be embedded in. We structure our environment into both intentional and non-intentional ingredients, but the former are not less "natural" for that, nor in need for "naturalization". The physical-intentional distinction has first of all a cognitive basis, not a metaphysical one, and in this sense we do not need normativity for there to be mental aspects in the things that surround us. We need a mind of the right sort. While that account questions the very basis of the naturalization project (which depends on setting up a *metaphysical* dichotomy), it also questions the idea of anti-naturalism, which depends on the same spurious dichotomies.

If mental and intentional aspects of nature (beliefs, desires, intentions, etc.) are simply there in our environments because we have the kinds of minds we have, the metaphysical idea of a "reduction" becomes neither affirmable nor deniable. It becomes simply obscure. Making the existence of beliefs and desires contingently dependent on evaluating them according to certain norms is no more invited here. What we are left with is a *methodological* (rather than

metaphysical) naturalism that inquires into the mental aspects of animate beings in no other ways than it inquires into their physical aspects. Studying human language is to study just another organic system with its internal structure and function, like the immune system or the circulatory system.

<div align="center">4. Pragmatism and naturalism</div>

Neither pragmatism nor rationalism, I take it, is generally speaking inconsistent with a methodological monism and naturalism of the above variety. Pragmatism wasn't in the case of Dewey, an issue to which I turn shortly. Rationalism certainly wasn't either, not for example in the case of Descartes, whose supposed "dualism", to the extent that this term makes sense (cf. Baker and Morris 2002), was formulated as a part of the natural sciences of his day. It certainly isn't either in the case of Chomsky's "rationalist psychology" (cf. Chomsky 2002), or the biolinguistic tradition (Lenneberg 1967).

Pragmatism met with naturalism particularly in the Dewey of *Experience and Nature* (1926).[10] Both Levi and Dewey's model of inquiry emphasizes the concept of equilibrium (cf. EK, section 1.5), but the specifically biological idea of (what we would today call) homeostasis is important only to Dewey, where it plays a role both in his analysis of life and in his epistemology (cognition being a homeostatic mechanism). States of organic equilibrium are disturbed, but then, through actions on the environment, equilibrium is restored. Thought on this picture is *among* the organic activities whose basic function is the restoration of equilibrium. Dewey's use of the latter notion reflects the close affinity and continuity Dewey saw between life and mind generally. A state of organic disequilibrium is a state of *need*, and one might say that a state of *doubt* is nothing but the mental version of this same kind of structural pattern. As Godfrey-Smith (1998, 105) puts it:

---

[10] See Godfrey-Smith (1998), Ch. 4, for a discussion of Dewey's naturalism.

"The disequilibria in living organisms induced by environmental events are like 'proto-problems' for Dewey, and all living activities which act to regain the organic equilibrium are proto-solutions".

This does not tie in well with Levi's normativist -- hence discontinuist -- picture of the mind. Inquiry in its most general conception is a response to an "indeterminate situation" for Dewey, where the indeterminacy is crucially *recognized* by the agent, not *created*. We are talking about an objective indeterminacy, and about inquiry as inducing an "objective change to the situation, transforming indeterminateness to determinateness" (Godfrey-Smith 1998, 113). That thought restores determinacy is an idea that might well, were it not for its naturalistic character, provide a foundation for Levi's idea that agents commit to the truth (and not just to maximal probability) after going through a situation of doubt (indeterminacy). The commitment to the truth would be the moment where a new determinate base of thinking (and potentially new problem-solving) is restored. Truth having no foundation in Levi (cf. EK, Ch. 1), and having no relational definition in terms of correspondence or ontology either, it may in fact amount to little more than determinacy: a stable state. It is a stable point from the perspective of which a new indeterminacy can arise and be dealt with.

I like this way of putting things, because so many have wondered how, on Levi's legendary infallibilism and absolutist conception of knowledge (EK, sect. 1.6), where the agent is committed to rule out the falsehood of any logical consequences of what he fully believes as being no serious possibility, such a state could ever be rationally either risked or revised. If knowledge is no more than a point of determinacy in the sense just recommended, the pressure of this puzzle seems resolved -- on naturalistic lines, or on the grounds of the fundamentally

practical needs of a natural creature. Few would doubt that an environmental problem posed to an animal -- catching prey, preserving body temperature, or high-feeding -- never gets solved. Such problems are not held open for all times. Why should cognition be different, and humans keep an open mind on any issue at all times, reckoning, as a matter of principle, *any* odd possibility a serious one? Should we expect natural selection to engineer *such* a creature?

Incurring a commitment can be to decide over truth and falsity, then, and the uncertainty-inducing, *relational* question of whether, in a given state of determinacy, we really have hit upon the truth, never arises. Settling the truth is what the state *consists in*. Comparing a given state of knowledge with reality is excluded almost by definition. Truth is no relational notion, and may be viewed either as a commitment to having dispositions to endorse certain consequences (those entailed by one's current state of full belief), or as a value that may be evoked when justifying a change of that state.

5. Competence and Commitment

Returning to human infants, we may say that they are naturally equipped with "systems of knowledge" about certain domains (language, physical space, social relations, mental states, etc.). Developmental biologists and psychologists speak of "innate expectations" about the structure of these domains, but the idea appears to be the same: infants come to encounter objects, persons, languages, etc., knowing a great deal about them.

Can we say as well that they encounter them *believing* a great deal about them? I see no motive for this particular move. The picture of the child forming particular beliefs or theories about some particular domain-specific problem seems a strange assimilation of a model of theory-formation possibly appropriate to science. Whatever the application of the apparatus of propositional attitudes in domains of adult problem-solving, it should be dispensed with if it yields no further explanatory benefit in the description of the child's language faculty as simply a

complex dynamical system in nature that undergoes certain state changes prior to maturing and stabilizing around the age of puberty. If language acquisition is basically a process of maturation that depends to a very minor extent on environmental feedback or cultural difference, and leads to success in all cases except severe pathology, there is no point applauding the child for its well-conducted "inquiry" once the feat is done. An *evaluation* of cognitive success according to principles of rationality appears misdirected, and the picture of the child as undergoing a "contract" when risking a new commitment in the light of its contingent values and goals, somewhat bizarre.[11] That does not mean that the terminology of belief and rational belief change yields no benefits in other respects, and I turn to these later on.

We have now moved from a commitment-performance distinction so crucial to Levi's picture to a competence-performance distinction. Performance is not as such subject to scientific study, as a vast set of cognitive faculties and background assumptions enters in even the simplest act of language use. Idealization and abstraction is the only way to study reality in its full

---

[11] If possessing grammatical competence (as opposed to chimpanzees or cats) is a biological feature of us having nothing to do with contingent values or rational performance in Levi's sense of these terms, it may be that our *logical* competence should be described in a similar way. Having a mind structured in whatever ways evolution structured it, we find us compelled accepting certain inferences as compelling. On this view, it would be wrong to talk about our (on Levi's view, unrevisable, cf. CR, 17) "beliefs in the truths of logic". We have no such beliefs, we just have a mind with structures that turn out usable in various ways. Related to that, I also disagree with Levi (2002, 125) that the notion of (linguistic) meaning becomes redundant once the notion of belief and the apparatus of rational belief change is adopted. A naturalistic attempt to explain, on the basis of linguistic principles, why a particular sound structure corresponds to a particular meaning structure, is untouched by work on the revision of our beliefs. Analyticity, too, as I have argued elsewhere (Hinzen in press, b), should not be discussed in either the epistemological or the alethic mode. The immediate cognitive steps from *I painted Bettina blue* to *The surface of Bettina's body is blue* (not Bettina the person), or from *My Saab has a Ford T engine* to *My Saab runs by means of a Ford T engine* (rather than having a Ford T engine stored in its trunk, which is a possible meaning of *There is a Ford T engine in my Saab*) are not inferences guided by norms for how we rationally change our minds. They have explanations in terms of how evolution and system-internal constraints have built our shared human faculties of language or linguistic competence (for the explanation of the last example, see Hinzen 2003). To the extent that these explanations are correct, there is no need to assume that these analytic truths have anything to do with belief, cognitive evaluation, or how we change our minds.

richness. With this, it seems, we must agree, and idealization and abstraction will not bring us outside the confines of naturalistic inquiry. Evaluating performance against a background of *commitments*, by contrast, will. It is with this methodological decision that I take issue, as it is this very step which deprives us of human rationality as a subject matter for naturalistic inquiry.

While having this sort of subject-matter is a crucial aspect of Cartesian and Chomskyan rationalism as I understand it, foundationalism of the sort that Levi's pragmatist axiomatically rejects, is foreign to it. While some anti-foundationalists today combat the "foundationalism" of the rationalists, there is a sense in which there wasn't ever any serious issue of foundationalism after the epistemological crisis of the late 17th and early 18th centuries.[12] But in a sense the entire Platonic rationalist conception of knowledge, as it transpires in the *Meno*, is anti-foundationalist. Here knowledge comes for free, resting on no evidential foundation in experience. It is natural for humans qua humans, irrespective of training or formal education, to have certain systems of knowledge. In the light of some experiential triggering, knowledge "dawns in us like in a dream", as the *Meno* (85c, 9-11) puts it strikingly, not a particularly rational way of coming to knowledge, one should think. How one can misread rationalism as inaugurated in the *Meno* and as carried much further in Descartes' philosophy of science as a "foundationalism" seems obscure.

Rationalism as I understand it, then, is the thesis that having certain forms of knowledge is part of having the nature of a human being. Human rationality is part of human nature, and is as much subject to empirical study as the latter is. A "rational psychology" is no more than a science of the mind based on the insight that human cognitive performance makes sense only in the light of certain posited systems of idealized competence, describable in terms of certain principles and rules that a human uses to analyze experience. Psychology is the study of knowledge, of which performance is only the kind of indirect evidence from which we have to

---

[12] I am indebted to correspondence with Noam Chomsky here.

take our clues. Whatever our use for a commitment-performance distinction, in the study of the mind the need for competence as an explanatory notion seems hard to escape.

6. Levi's Principle

What I called "Levi's principle" in Hinzen (2000), according to which "deliberation crowds out prediction" (CR iv, 31-32, 76-79), brings an interesting twist into the story told so far. Evidence for this principle is perfectly intuitive, on the assumption that principles of rational choice have their intended application in the agent's monitoring the rationality of his own decisions. For such efforts are vacuous if the principle fails: principles of rationality are meant to apply *within* deliberation and help an agent to reduce a set of options judged to be *feasible* to a subset of *admissible* ones. If within deliberation a choice is predicted, say on the basis that the choice will be *rational* or that an *admissible* option will be chosen, the choice is vacuous. For the agent should thereby become certain that the option is chosen, and then no other choice is, as far as he is concerned, feasible.[13]

That said, note that in principle the opposite conclusion could be drawn from the pre-emption result just reached: principles of rationality are not to be used in the agent's monitoring of his own decisions but in other agents' predictions and explanations of his actions, or the theorists' (Levi concedes as much, cf. CR 26, 31, 35). That is, rather than concluding, from the use of rationality principles in deliberation, that they cannot serve for prediction, we now conclude, from their use in prediction, that they cannot be useful in deliberation. Indeed, if human actions are parts of the natural world, and this world is inherently probabilistic, why should

---

[13] In fact, Levi's principle forbids that the agent assigns *any* probabilities on his potential acts, not just the extreme ones (CR, 32, 76-7). Rabinowicz (2002) argues that this strong version is based on a too tight connection between probabilities and betting rates. Hinzen (2000) points out that the rather far-ranging consequences of Levi's principle in this strong form with respect to game theory and the explanatory power of its various solution concepts should be weighted against the clear use that these solution concepts appear to have, particularly when placed in an evolutionary (i.e., non-normative) setting.

human actions not be probabilistically predictable, whatever our first-person perspective on the matter? Of course, if principles of rationality are false as applied to human beings and thus not predictive, this is not an attractive line to take. But then, this is not an objection, for maybe one should *abandon* rationality theory as a theory of human behaviour, rather than *re-interpreting* it in normative terms.[14]

On Levi's own view, if principles of rationality were applied to rational beings (which, to repeat, he thinks we are not), they *would* be predictive: a sequence of actions will follow deterministically from them. But they aren't, as a matter of empirical fact, so we must not assume we are rational beings, as long as we stick to our principles of rationality. We are committed not to assume that we will act rationally (the "smugness assumption", CR, 31). Our necessary uncertainty in our rationality is the price we pay for having principles of rationality making, as is natural for such principles to make, predictions on the behaviour of rational beings. But this feels like going a step too far: being told one acts irrationally seems bad enough, but being told one must necessarily not assume to act rationally, or to assume one's own confusion (else one could predict one's actions, hence could not choose) seems somewhat harder to swallow. Again, why save the theory of rationality from being a refuted theory by stipulating a new prescriptive principle forbidding its use in prediction?

We need a motive for Levi's principle other than the one that it, if not accepted, makes a given theory have wrong predictions. That principles of rationality make predictions for beings

---

[14] Even given such a re-interpretation, it is not clear to me why a deliberating agent should not be allowed to assess the rationality of his acting in accordance with what principles of rationality predict: while deliberating, he might switch to another mode of thought, pausing to wonder how rational he will likely choose today, given his past "record" of acting rationally. While these are indeed two different modes of thought, which would lend support to Levi's view, I can't see why it is rationality (rather than mere psychology) that prevents him from entertaining these two kinds of thought at the same time. This line of thought suggests that Rabinowicz's (2002, 92-3) "screening off" suggestion for justifying Levi's principle won't do: it points to a merely psychological incapacity, whereas Levi intends his principle prescriptively.

that are described by these principles is after all the first and natural thing to assume. Why on earth, if they apply to us, should they not be predictive? Well, we mentioned the standard reasons offered: limitations of computational capacity, for example. But one can accept these limitations without giving up on the descriptive status of principles of rationality and going for Levi's strictly normative model instead. In fact, the way that systems of competence interact with systems of performance using these systems of competence provides a different account of the same phenomenon, without normativity entering. Thus, if we assume, with tradition, that our language faculty can be described as a combinatorial system consisting of primitives and rules -- a grammar -- then we find in this domain as well that there is "no end to human confusion and diversion" leading humans to produce ungrammatical sentences. Grammatical rules being recursive, for example, they allow us to construct ever longer and longer sentences. Memory limitations will eventually disallow us from processing them, but that will not as such be an objection to them being generated according to the rules and primitives of the grammar. The grammar system, we will conclude, interacts with other systems, leading our theory of grammatical competence to make wrong predictions in an infinity of cases. It will be explanatory all the same, for we will appeal to the rules of grammar as entering into the complexities of language use and as explaining an aspect of their full richness. No motivation for normativism arises, and the question looms large of why we should not go for a similar account in the case of the human "faculty of reasoning" (to use a probably very misleading term), rather than prevent predictive and explanatory uses by additional stipulations.

In a discussion motivating his principle, Levi (CR, 37) argues that norms of rationality are also not "blueprints for rational automata". That would exclude robots from the intended users of norms of rationality, but again I lack faith in the distinction made here, which seems to reflect a non-naturalistic bias and dichotomy: humans deliberate (use standards of rational health for self-

criticism), robots don't. The point appears to be that rational automata function according to deterministic laws, so that laws of rationality applied to them would lead to predictions of rational choices. But then again, why not let principles of rational choice have their predictive uses, while arguing that, in the case of humans, the system described by these principles are embedded in a quite different cognitive architecture than it is in robots, leading to interaction effects naturally excluding predictive uses of the same kind?

Robots are different from humans, to be sure, but a naturalist will start by assuming these differences are empirical, not categorical: unless we beg the question for anti-naturalism, it is not that we are also, but not only, physical beings, while robots are only physical beings. Assuming naturalism, we cannot make the empirical differences depend on common sense intuitions we have concerning ordinary words such as "choice" or "deliberation". For native speakers of English and probably all other languages "choice" as well as "deliberation" analytically contradicts "deterministic prediction". Levi's principle takes its intuitive appeal from right here. But then, it is rational systems we wish to study -- human beings -- not the meanings or conceptual contents of common sense terms. If the robot does not "deliberate", do we? Of course. But from another point of view, such concessions lead to no comfort. Deliberation is a process that can be functionally described. We are, for all we can tell, lumbering robots, even though we like to describe us differently, and no doubt are well-advised to do so in our daily lives.

7. Finale: Rational beliefs in a naturalistic perspective

As Jon Elster (1989) emphasizes in his discussion of rational choice, explanatory weakness is not at all, in general, a necessary consequence of predictive weakness. A mechanism may easily have explanatory uses, but not allow us to make predictions, precisely because it may interact with a number of other mechanisms which may disturb the functioning of the first. In other words, we may not be able to tell when one among a number of possible explanatory mechanisms is

enacted. Taking this into account, it seems that although being interpretable as rational may not be *constitutive* for being in a mental state, it may sometimes be *illuminating* to appeal, in describing the workings of the mind, to the operating of a *mechanism of rational choice.*

If that mechanism operates, the person's behaviour will be optimally adapted to the given circumstances. In Elster's picture, that optimality is internally differentiated in three distinct optimizations: the person has optimized the time to spend on the collection of evidence; the person has formed the optimal beliefs given the evidence (rather than merely being led by wishful thinking, say); the person has determined the optimal means for realizing his desires given his beliefs. The role of beliefs is essential here. Being in a state of belief means to look at the world in a way that my perspective is not coloured by features that the world would have only if it were as I desire it to be. I imagine this like a filtering process in which only those features remain that are not induced by my desires. It is clear that only specific productions of the language module could be usable by a belief-producing mechanism thus understood, not for example *I'll leap in through the window*, which is not my mind's construction of a fact, but is a projecting of a behaviour of mine into future developments of the present state of affairs that are not yet factual.

Despite the fact that the rational choice mechanism points to a sheer luxury that nature affords (it's just fantastic to ever get so optimally adapted), there seem to be rather clear cases where the rational choice mechanism has a role in the explanation of behaviour. In those cases, postulating rational beliefs as theoretical entities in a naturalistic account that play a role in the chemistry that produces an action does not seem far-fetched at all. There are also other types of behaviour, which, although they are not rational in the full sense, involve beliefs in a naturalistic sense of the word. Take for example the mechanism in which the belief that I cannot get something that I desire causes me to stop desiring it (the so-called sour-grapes mechanism).

Still, the rational choice mechanism will not serve as a general theory of behaviour. There are specifiable conditions where mechanisms are enacted that give no role to rational choice as an explanatory mechanism. As Albin (1998) shows, some decision problems have no rational solution for reasons of computational undecidability in the Goedelian sense, and also cannot be known not to have such solutions. In such cases, postulating rational expectations or the formation of rational beliefs seems vacuous. More down to earth, it may be a sheer *lack of opportunities* that dictates and explains a person's action. The economy of those actions will do without the mechanism of rational belief as well. For a last example, action explanations often appeal to the mechanism of weakness of the will, or the fact that although I desire something strongly, there are other desires that win over, simply because they develop a powerful psychic turbulence. In this explanatory mechanism beliefs again play no clear role.

A coherent view then may be this: the human mind is a toolbox of mechanisms of different types. In some, rational beliefs do play a role, in others they don't. If so, the category of belief will have to be investigated much more carefully with respect to the different ways in which the mind enacts mechanisms for choice.

We began with an analysis of Levi's normative conception of rationality, and expressed our desire, not only to view human behaviour and its explanations as part of the natural world, but also to see principles describing our rationality as explaining why we act in the way we do. Levi's objections to the ordinary notion of belief's having a naturalistic correlate and entering the explanation of human action will not convince a methodological naturalist, not least in the light of recent findings in cognitive and developmental psychology. No clear case emerges for necessarily analysing human rationality as intrinsically emanating from a set of commitments, as opposed to a number of cognitive and interacting competences that naturally belong to us as humans. If mental aspects belong to us like other organismic aspects do, a recommendation on

what well-conducted inquiry or change of mind is should be paired with a study of the mind and the intrinsic structures that make it up: the mind we do not change.

References

Albin, P.S. *Barriers and Bounds of Rationality*, Princeton University Press, 1998.

Baker, G., and K. J. Morris. *Descartes' Dualism*, Routledge, 2002.

Bjorklund, D. F., and A. D. Pellegrini. *The Origins of Human Nature: evolutionary developmental biology*, Washington, 2002.

Chomsky, N. *On Nature and Language*, eds. A. Belletti and L. Rizzi, Cambridge University Press, 2002.

Conlisk, J., "Why Bounded Rationality?", *Journal of Economic Literature* 34, 2 (1996): 669-700.

Cosmides, L., J. Tooby *The Adapted Mind*, Oxford University Press, 1992.

Dennett, D. *Darwin's Dangerous Idea*, Penguin Press, 1995.

Elster, J. *Nuts and Bolts for the Social Sciences*, Cambridge University Press, 1989.

Foley, D. K., "Rationality and Ideology in Economics", Ms., Department of Economics, New School University, 2003.

Godfrey-Smith, P. *Complexity and the Function of Mind in Nature*, Cambridge University Press, 1998.

Harman, G., "Moral Philosophy and Linguistics", Ms., Princeton University, dated November 11, 1998.

Hinzen, W., Review of Levi (1997), *Erkenntnis* 52 (2000): 403-407.

Hinzen, W., "Truth's Fabric", *Mind and Language* 18, 2 (2003): 194-219.

Hinzen, W. *Minimal Mind Design*, two volumes, Oxford University Press, 2004, in press, a.

Hinzen, W., "Der gegenwaertige Stand der Gebrauchstheorie der Bedeutung", in *Pragmatisch Denken*, eds. Fuhrmann, A. and E. Olsson, ontos-Verlag, 2004, in press, b.

Hirschfeld, L. and S. Gelman, eds. *Mapping the Mind*, Cambridge University Press, 1994.

Hume, D. *A Treatise of Human Nature* (1739-40), ed. Selby-Bigge, L.A., Oxford: Clarendon 1978.

Kuhlmeier, V., Bloom, P., and Wynn, K., "Do infants see humans as material objects?", *Cognition*, in press.

Lenneberg, E. *Biological Foundations of Language*, New York: Wiley.

Levi, I. *The Enterprise of Knowledge*, Cambridge: MA: MIT Press, 1980.

Levi, I. *The Fixation of Belief and its Undoing*, Cambridge University Press, 1991.

Levi, I. *The Covenant of Reason*, Cambridge University Press, 1997.

Levi, I., "Seeking Truth", in Hinzen, W., and H. Rott, ed. *Belief and Meaning - Essays at the Interface*, Frankfurt: Haensel-Hohenhausen, 2003: 119-138.

Mehler, J., and E. Dupoux. *What Infants Know*, Blackwell: Oxford, 1994.

Pinker, S. *The Blank Slate. The Modern Denial of Human Nature*, Penguin Books, 2003.

Rabinowicz, W., "Does practical deliberation crowd out self-prediction?", *Erkenntnis* 57, 2 (2002), 91-122.

Rorty, R., "The Priority of Democracy over Philosophy", in Malachowski Alan R., ed. *Reading Rorty*, Oxford: Blackwell, 1990: 279-302.

Wertheimer, R., "Philosophy on Humanity", in R. L. Perkins, ed. *Abortion: Pro and Con*, Cambridge, MA: Schenkman, 1974.