



A Non-Substantialist Approach to Practical Reason

by **Yujian Zheng**

Lingnan University
Tuen Mun, Hong Kong

IJFB, Vol. 2(1) Article No. 6, 2002.

Citation URL: <http://www.iifb.org/ijfb/YZheng-3-6>

[Back to Table of Contents](#)

[Format to Print](#)

- ¶1. In this paper, I shall discuss in section I. the conceptual ingredients of being irrational, viz., its relation to reason and to non-reason. In sections II. and III. I shall deal directly with the possible paradoxical elements in explaining the irrational, and discuss certain notions and distinctions concerning mind and its interpretation with an eye to explaining away those *prima facie* paradoxical elements. Section IV. will delineate a way to transform Davidson's paradox of irrationality into a new form of methodological challenge that explicitly draws on the distinction and the interaction between descriptive and normative factors in rational thought and is likely to generate other interesting practical questions. And finally, in section V., I shall show the weakness of Davidson's expression of the paradox of irrationality by revealing two intuitive assumptions underlying the expression. My conclusion then about the non-substantialist, normative nature of human practical reason or interpretive rationality will be directly derived from these considerations.

I.

- ¶2. Typical irrationality can be exemplified by akratic or incontinent actions, i. e., free or intentional actions against the agent's best judgments. Although there has been no consensus on the diagnosis of what really happens in such actions, people seem to all agree to call them "irrational." Among various accounts of what irrationality amounts to, Donald Davidson's view stands out conspicuously.
- ¶3. Irrationality is a failure within the house of reason, says Davidson.^[1] That is, if a person never engages in relevant reasoning, he cannot be judged irrational; we may call him nonrational at most. Though the irrational must not lie totally outside the ambit of the rational, it must not lie totally inside it, either. "Totally inside" means that everything involved in the putatively irrational behavior has an explicit or implicit reason, the explanation of which is so good as to turn it into one or another form of hidden rationality; or, in other words, the actor himself may, but does not necessarily at the moment of action, appeal to reasons and explanations to justify his apparently "irrational" behavior. "What is special in incontinence is that the actor cannot understand himself: he recognizes, in his own intentional behavior, something essentially surd."^[2] This challenges us to explain a certain kind of putatively irrational behavior without thereby rendering that very behavior subjectively rational—i. e., from the actor's own point of view, his action is irrational, no matter at the moment of acting or later when he has more time to reflect upon it. So something must be partially outside the ambit of reason that has done the job of making irrationality. In other words, the irrational acts must involve both a rational element (viz., appealing to reasons) and a nonrational element (viz., being susceptible to non-reasons) at the core; short of either element we will not have irrationality. Davidson seems to follow this line of thought when he begins to appeal to a partitioning model of mind and claims that reason has no jurisdiction across the boundary.^[3] Interestingly enough, however, many theorists have tried to point out that Davidson's treatment of incontinent actions in reconciliation with his principles concerning the relationship between reasoning and acting is unsuccessful, even on his own terms,^[4] and too intellectualistic, viz., focusing too much on the cognitive powers such as reasons and judgments and neglecting the role of conative powers in motivating such actions.^[5] Although it is important to see that irrationality, if there is any, only occurs when conative factors come into relations with reason, i. e., only when it is not a pure matter of (cognitive) reasoning, none of the above criticism has ever touched one important notion of the paradox of irrationality, initiated by Davidson:

The underlying paradox of irrationality, from which no theory can entirely escape, is this: if we explain it too well, we turn it into a

concealed form of rationality; while if we assign incoherence too glibly, we merely compromise our ability to diagnose irrationality by withdrawing the background of rationality needed to justify any diagnosis at all.[\[6\]](#)

Many theorists who try to provide an adequate explanation of weakness of will and its bearing on the issue of rationality fail fully to appreciate the implication of the above remark, something that I believe is potentially significant for any attempt to understand the source of many apparent puzzles around irrational behavior such as weakness of will. This failure may be partly due to the fact that Davidson himself does not make it clear, or even is not fully aware of, how central this thesis may be in his own account of action and its interpretation.

II.

- ¶4. We can rehearse the Davidsonian paradox as follows. If we explain too well, for instance, how weakness of will is possible, i.e., if we find a causally and mentally too coherent picture for it, we turn it into a concealed form of rationality; while if we assign incoherence too glibly, i.e., if we exclude weakness of will from the class of intelligible phenomena (for incoherence breeds unintelligibility), then we merely mystify the otherwise understandable (thus rationally diagnosable) human behavior by giving up the always possible effort of successful interpretation of the other. Put in this way, it sounds more like some hermeneutic dilemma. But sceptics may question the genuineness of this kind of paradox. What is the demarcation between explaining appropriately and explaining "too well?" If coherence is the objective of our interpretation and explanation, how can it be criticized as too coherent? On the other hand, if someone fails to understand a difficult or complicated phenomenon because of his laziness or incompetence yet excuses himself by blaming the object for incoherence, that is not puzzling at all and would be just an object of censure. If we have no doubt about the fundamental coherence of nature, why not stop assigning any incoherence to human nature as well? Then, what you are left to do is try your best to get as coherent a causal explanation as possible for any human phenomenon.

- ¶5. Our belief in the ultimate coherence of nature provides the basis for our pursuit of explanatory coherence. One implication of natural coherence for any explanation is that every item in the explanandum has to have a natural cause which is sufficient for its occurrence and all natural causes operate in ways that are consistent with all natural laws. But it is not the case that explanatory coherence can automatically warrant any form of rationality on the part of the agent whose actions are being explained. When Davidson worries about turning an irrational action into "a concealed form of rationality" by explaining it "too well," he seems to wave his hand at a potentially problematic complex of factors, the net effect of which tends to lead us astray in our rational assessment of the action. But Davidson fails, for whatever reason one might speculate, to offer any analysis of such a complex. What I will do here is try to offer an analysis to the effect that one can make optimal sense of whatever insight or rationale underlies that particular remark of Davidson and this will put us in a better position to examine the validity and significance of the alleged paradox.
- ¶6. In order for any complaint against "a concealed form of rationality" to get off the ground at all, it seems to me that one needs first to offer a plausible account of how a causal explanation can directly lead or relate to rationality ascription under normal or standard situations; and this, hopefully, will yield a needed background principle that a good causal explanation, say, of an appropriate type can legitimately turn the behavior being explained into a form of rationality. Only after this picture is well placed could one start talking about possible scenarios in which causal explanations could go astray, including being described "too well," so as illegitimately to turn the behavior into a certain form of rationality. Think in another way: suppose that under no circumstances causal explanations of any kind have anything directly to do with rationalization; then why would anyone on earth worry that certain particular causal explanations might illegitimately rationalize some presumably irrational behavior? All one could then say is this: this behavior is shown by this explanation as causally coherent but rationally irrelevant or undetermined. On the other hand, however, the stronger the link between causation and rationalization either generally or to a significant extent, the more real the worry about the possible mismatch between the two, or about possible mistakes involved in identifying or revealing the genuine causal items in the explanation.

- ¶7. So it seems that making sense of the alleged paradox depends upon a clear account of the relationship between explanatory reasons (causation) and justificatory reasons (rationalization); and such an account in turn depends upon an appropriate formulation of a possible background principle mentioned above.
- ¶8. Let us start with an extreme version of the principle: i.e., "in the domain of human actions (including mental acts), any causal antecedent can rationalize (viz., serve as a justificatory reason for) its consequent." Apparently this is hardly acceptable simply because of the fact that some causal items or elements in that domain obviously have nothing whatsoever to do with reasons or reason-giving relations and, therefore, rationalization is out of the question for them.
- ¶9. Now let's limit the "causal antecedent" in the above version to "causally explanatory reasons." Does the resulting version, i.e., "any (causally) explanatory reason for an action (which follows as the causal consequent) can rationalize it," look more tenable? The answer seems positive - at least when we limit the sense of "rationalizing" to that of an instrumental satisfaction (i.e., the action is a reliable means to the end which constitutes the reason) in a separable, local perspective (viz., without comparison with other available or potential reasons). Here, however, is not the place to offer or examine any detailed arguments for this version of the principle in various action-situations. For the sake of our mainline argument, let's grant the validity of this principle.
- ¶10. When Davidson worries about turning an irrational action into a concealed form of rationality, he seems to suggest a possibility that the explainer may put some imaginary item such as a causal reason into the action (as "hidden variables," so to speak) in order to make the action look coherent in some easy and artificial way. Given the above principle, such an artificial move would in a certain sense rationalize the action. The question, however, is how anyone can be sure of one's judgment that a certain item or variable is imaginary rather than really operative in some unrecognized way.

- ¶11. Although such a question can be raised for any natural phenomenon to be explained, it becomes more important when one faces a human phenomenon which involves psychological properties of various kinds. A psychological event can be both causal and mental at the same time, meaning that it can be described in both neurophysiological (and ultimately physical) terms and in terms of its propositional content or logical relations to other events with propositional contents. But before we can get any detailed picture of the relationship between mental and non-mental/causal terms, there is much conceptual work to do. What is special about mind is not just that it creates and understands meanings, i.e., has certain mental states or events with propositional contents, but also that the mental states, described in folk psychological terms, such as belief, desire, reason, judgment, want, will, volition, emotion, etc., do not have clear conceptual boundaries. They normally overlap, interweave and change into one another. What this implies for the issue of diagnosing irrationality is that we usually have enough maneuvering room for either explaining (rationalizing) or dismissing (ruling out) puzzling phenomena as we find it exigent to do so. Such maneuvering room or freedom in our interpretations is made possible by lacunas, overlaps and fuzzy or gray areas among these conceptualized mental items in linguistic forms.
- ¶12. Precisely because of this kind of unavoidable conceptual looseness, we always have a choice between assigning this or that level of coherence to this or that part of mind with corresponding propositional contents or logical relations. This implies that there is normally more than one way to obtain the targeted holistic explanatory coherence, and it is thus conceivable that a certain way of obtaining holistic coherence could better (than other ways) preserve the integrity of the normative rules we accept as our basic principles of reasoning and/or action.^[7] When we rule out an assumed mental item in an action as unreal and base our criticism on the breach thus "created" on the part of the agent, we are often not doing it arbitrarily but rather following the relevant rules as far as possible. Even when we talk about easy and natural behavior under normal circumstances, we are also following rules of various kinds; and the rules we follow or the ways of applying them could be contextually wrong without our recognition. Hence there is *always* an issue in mapping out the mind under explanation in a certain way, i.e., whether we have struck the best balance between fitting individual mental items with the overall observable behavioral patterns on the one hand, and keeping our critical ability in following certain fundamental principles, which constitute our contemporary horizon of rationality, on the other.

- ¶13. But where exactly does the paradox lie except in the tension embedded in the above issue of striking the best balance?

III.

- ¶14. The degree of rationality with which a person behaves, according to common sense, seems to be some inherent property of his, which can thus never be dependent upon the third-person view of his behavior. At least at any particular moment of his life, it seems to be a fixed fact about his mental capacities or tendencies that he exhibits certain general patterns of thinking and motivation. Without disputing this, how can we then make sense of the claim that the interpreter is entitled to assign a certain degree of coherence/rationality to the person being interpreted?^[8] In other words, is it not paradoxical to say that some independent fact about one mind is dependent upon another mind's decision to, or how to, understand it?

- ¶15. One reply to this question is perhaps that this apparent paradox can be easily eliminated by a more careful form of expression that draws on the distinction between ontological independence and epistemological independence: i.e., the ontological independence of a mind, or any object, from the interpreting subject does not contradict the epistemological dependence of an action upon the conceptual apparatus and interpretive strategies of the subject. So one can consistently say that any "factual" statement about an action is nothing more than a descriptive form of understanding from a certain interpreter and the interpreter need not deny the ontological independence as well as coherence of the action. Even the dual character of mental events, viz., being simultaneously causal and propositional, cannot really change the validity of this distinction, unless one can demonstrate that some particular interpretive moves of the subject tend to affect the ontological status of the action.

- ¶16. Specifically, the ontological assumptions required by this distinction are that any natural thing or event is coherent and thus mind must be coherent in virtue of its being a natural thing. Such assumptions are independent of or prior to any particular human understandings of mind. When we judge some action internally irrational, i.e., the actor intentionally violates the criteria he has freely adopted and is presently holding, are we ascribing some natural incoherence to the actor as if it were a mysterious defect in his natural makeup? If the answer is positive, a problem of misconception seems to have occurred: we transgress the bounds of ontological assumptions with reasons that merely have epistemological relevance. If, on the other hand, we contend that what we are ascribing is not ontological

incoherence, but rather explanatory incoherence, viz., the incoherence in *our* explanatory scheme between a certain supposed or imagined mental item of the actor and his other more or less well-established mental parts as well as certain well-accepted principles of action/thought, then the question is why we should ascribe to the actor such an interpretive problem, which seems only to reflect some imperfect or incompetent state of ourselves as interpreters? And a further question will be, "what purpose does it achieve to judge certain actions internally irrational rather than temporarily unintelligible?"

- ¶17. There does seem to be a certain oddity in attributing irrationality on the grounds of some incoherence which is either groundless at the ontological level or misplaced from the part of the interpreter to that of the actor. Real incoherence can only exist between mental states not as pure causal items but rather as reasons or propositions-this must be true for both interpreters and actors. No matter what other possible sources of the explanatory incoherence there are on the part of an interpreter, there is always the possibility that it is caused by, or corresponds to some real incoherence on the part of the actor between his own reasons for action and other underlying (perhaps unknown) mechanisms of action - also expressible in propositions. Therefore it is always open to interpreters as a default or alternative strategy to ascribe the origin of whatever explanatory residue to the actor side. Where or when they are fully justified to do so is an empirical question that depends on many contextual factors as well as certain general normative concerns such as those expressed by "the principle of charity."[\[9\]](#)
- ¶18. One relevant point, however, needs clarification here. Certain instances of incoherence between the actor's own reasons or propositions need be nothing but results of simple mistakes, such as errors in calculation, caused by absentmindedness, forgetfulness, or insufficient attention, etc. These incoherent thoughts cannot count as instances of irrationality, for they are not intentional violations of any criteria the actor endorses. Therefore, interpreters must be careful to avoid premature accusations of irrationality, i.e., to distinguish such shallow sources of incoherence from other deep sources of incoherence on the part of the actor, for only the latter could provide the possible ground for the criticism of irrationality. When I say "shallow," I mean that it is relatively easy to eliminate or circumvent the sources once they are identified; correspondingly, the so-called "deep sources" refer to those that are either unidentified or hard to eliminate. But what are these possible deep sources of incoherence that could give rise to irrational acts in a deep or strict sense? The question is hard to answer without appealing to empirical studies of mind such as cognitive psychology; what philosophers can offer, however, is a conceptualization of the types of relations between psychological

mechanisms and propositional contents. What seems puzzling in the case of internal irrationality is the apparent breakdown of normal reasoned explanations of the action concerned when reasoned explanations of intentional actions are among our basic methodological principles that are believed to be universally valid.

- ¶19. Of course, we always have the option of denying the existence of internal irrationality and we might make this denial a basic principle. The *Plato Principle*, named by Davidson,[\[10\]](#) does exactly this: it says that no one willingly acts counter to what he knows to be best. Can this principle really hold in the face of so many apparently incontinent acts in our daily life? One might argue that it would be too strong or unrealistic to sanction such a principle in the practical (rather than the theoretical) domain, which characteristically involves various conative and volitional factors.[\[11\]](#) But one may wonder why conation or volition would make such a difference. [\[12\]](#) It is interesting to notice that the above expression of the Plato Principle sounds like a pure *description* of a natural fact, i.e., everybody by nature acts, or reasons intentionally according to what he knows to be best or correct. But this could hardly be right unless you would define "what he knows to be best" here as nothing but whatever momentary state of his mind revealed by his subsequent, ostensible act, and thus make this statement trivially true, i.e., true by definition.
- ¶20. So the Plato Principle, although it looks like a descriptive truth, must be a *normative* principle whose proper function is not to describe or predict, but to prescribe or direct. In this light, we can make good sense of the point of judging an incontinent action as irrational: because it does not live up to the normative standard which the actor himself has naturally adopted and cannot rationally reject and because we believe that the action can be rectified according to that standard. Still, however, we never suppose that the irrational action is ontologically incoherent and thus we always keep the possibility open that one day we may be able to fully explain the natural genesis of the action in its utmost details; even then we are not "turn it into a concealed form of rationality," for we will still hold it as normatively incoherent, unless we somehow modify or abandon our normative principles in accordance with whatever new-found psychological truths might emerge. In a nutshell, as long as we are fully aware of the normative nature of these basic principles or requirements in our justifiable human practice, the original Davidson's paradox of irrationality seems to dissipate.[\[13\]](#)

IV.

- ¶21. A residual sense of doubt, nevertheless, seems to persist in the minds of those who are accustomed to take these basic principles more seriously than those maxims or rules that are obviously normative. Even after one is convinced of the normative basis of, say, the Plato Principle, one may still wonder why we should adopt this rather than that particular principle, with its particular boundaries of reference as well as application, as one of our normative principles of interpretation of intentional actions. Are there any natural, ultimate constraints on our freedom to enact and sanction such a normative principle? Should the content of the principle be constrained by the actual existing states of the minds whose actions are intended to be governed by the principle? In other words, should a normative principle demand descriptive accuracy of its actions or events? We can easily think of some normative rules whose sole purposes are but to demand conformity of behavior, for instance, a dress code for a workplace. The rule-makers need not pay much attention to whether, or to what extent, the people involved already have those behaviors required of them by the rules; the only concern here is that the required behaviors obtain as the result of complying with the rules. Many moral rules belong to this category. Obviously, there is no question of descriptive accuracy of these rules about the *status quo* of the actions or events in question.
- ¶22. There is a sense of paradox, however, when one talks about a normative principle of interpretation: for interpretation presumably targets some existing and fixed phenomenon and is thus truth-oriented and descriptive at its core, whereas normativity concerns solely the guiding standards we have adopted or agreed upon and is thus compliance-oriented and prescriptive. How can one reconcile the two opposite requirements in one principle without courting self-contradiction?
- ¶23. The key to a possible answer of this question lies in a crucial feature of the objects of interpretation: the actions being interpreted are not immune from the actor's beliefs, attitudes and behavioral responses with regard to the very principle by which the interpreters regulate their interpretive operations. We cannot ignore the possible effects of self-fulfilling beliefs, viz., the possibility that when people, including both interpreters and actors, sincerely believe in a certain principle, regardless of its ultimate descriptive accuracy, their behaviors may be so influenced by it as to verify its truth and reinforce their original belief. What makes the interpretations of human actions distinct from those of pure natural objects is that the interpreters must take into account the factor of possible interactions between the beliefs of actors about a principle and the believability or descriptive accuracy of that same principle which is applicable to the actors. Thus the validity claim of the principle cannot be limited to descriptive accuracy about the existing states of the actors in

isolation from their making attempts to adjust themselves to the principle, but should rather be inclusive of the possible states which the actors are capable of attaining - once they are determined to exercise their willpower to change the elements inside their minds that are responsible for their "deviant" actions. The question, however, is how the actors come to believe that those of their actions which apparently deviate from the principle are due to internal defects of some sort that need rectification. Criticism of "subjective irrationality," unlike other types of criticism, would be valid only if the actors themselves admit that their actions are outcomes of some defect which is somehow rooted in themselves (viz., not from some alien forces) and yet not excusable since no rationalization can be found for the actions.

- ¶24. Again, look more closely at the expression of the Plato Principle, "no one intentionally acts against what he knows to be best," or that of the utility principle, "everyone maximizes his perceived overall utility." They not only sound like descriptive statements, compared with the statements "everyone ought to act according to what he knows to be best" and "everyone ought to maximize his perceived overall utility," which fit *prima facie* with our ordinary patterns of intentional actions; but they are also able to survive ordinary skeptical scrutiny for their descriptive accuracy by the ineluctable conceptual maneuverability or variability of their terms, as discussed earlier. Perhaps this descriptive approximation plus the difficulty of finding reasoned explanations (viz., rationalization) for their "deviant" actions combine to offer some edge for convincing the actors who have troubles in squaring their own actions with the principles that their allegedly incontinent actions are due to some internal problems of their own, e.g., inconsistent thoughts that may or may not be testable by simple introspection; and that these problems, nevertheless, must be overcome by trying every possible means.
- ¶25. If this line of reasoning is valid, at least to a certain extent, then we are able to elucidate the real problematics underlying the original Davidsonian paradox of irrationality by transforming it into the following form of a methodological challenge: how can we actually reconcile the two seemingly legitimate yet directionally contrary requirements, viz., the requirement of objective description of the factual states of the action vis-a-vis the requirement of rational prescription for the behaviors of the event, in one and the same principle which is supposed to be either descriptive or normative but not both? From the above discussion, one can see that the normative intent of the principle is to be fulfilled, not by any direct, externally imposed mechanism of motivation through, say, rewards/punishments, but rather by some indirect, self-generated mechanism of motivation through the actor's own cognition of an internal problem which, as "problem" in a subjectively evaluative sense, is at least partially

constituted by his belief that he has deviated from something natural. This natural thing, however, is expressed by the very principle which, despite its normative nature, takes the form of a descriptive truth; and such a form is likely to be conducive to generating the actor's belief mentioned above. Does this mean that there must be a deliberate purpose under the appearance of the principle to mislead people about the genuine nature of the principle in order to achieve the normative end of realizing the very content it seems to describe? My answer is negative. For I can conceive of a world in which there is no such cunning purpose in any individual consciousness but the principle works along a similar path, i.e., the statement actually has a normative function of rectifying people's behaviors without any public action taken at a particular time to enact or reenact it as a normative principle. Of course, this does not suffice to deny that from time to time there may be individuals who can, or believe themselves able to, see through the misleading appearance of the principle and yet do not want to give away the secret for one reason or another. Put bluntly, the cunning reason, if there were any, behind human practice could work its way out with or without human understanding of its cause.

- ¶26. The remaining question, however, is whether the clear-eyed awareness of the normativity, whatever its origin, of a principle with the above character is not practically self-defeating, i.e., the more you are aware of the "secret," so to speak, of the principle, the more you find it untenable as a natural truth and hence the less effect it will have on guiding your efforts to live up to it. This is clearly an empirical question, which, fortunately or not, cannot be decided in accordance with our wishes. If the answer is yes, and we do not want to see the principle fail, then our best strategy will be an obscurantist policy, viz., to keep people in ignorance of its normativity as far as possible. Now there does seem to be a dual sense of paradox, however, if the "people" in the last sentence refers to, or includes, *ourselves*. On the one hand, this is a typical case of self-deception and many find it paradoxical to say that one can somehow believe the opposite of what is already known to oneself.^[14] On the other hand, the paradox appears as a matter of rule: the normative goal of a certain principle can only be realized when the people involved never regard it as normative; in other words, as soon as one becomes aware of the normative nature of a certain principle, its normative effects will disappear.

- ¶27. We can use a hypothetical scenario to show another implication of the alleged paradox here. Suppose that someone was once convinced of the normativity of a certain principle but after some skillful and successful self-deception, without his denying the irrationality of self-deception in and of itself, he now believes that the principle is a descriptive law and thus, by the motivational mechanism delineated above, he activates enough willpower to so regulate himself as to live up to the principle. Suppose again that we all (perhaps except for him) agree on the rationality of the principle as a *normative* one. Now what does this scenario imply if all the above hypothetical premises hold? The implication is that some rational goals are only realizable by some intrinsically irrational means. [15] For those who do not want to deny the proximity of this scenario to the reality of life, that implication may pose a new problem, whether you call it a paradox or not.

V.

- ¶28. Before I pull together the threads of the argument so far, I would like to explore on a deeper level the conceptual underpinnings of Davidson's original expression of the paradox of irrationality. What underlies Davidson's expression of the paradox seems to be the following two basic intuitions or assumptions. First, causal explanation of intentional actions can normally be applied to rationalize the actions in a minimal sense, i.e., not all of one's felt reasons for action are real or potential causes of one's action, but most identifiable causal items and/or motivational mechanisms an individual is possessed of will sooner or later, in one form or another, enter into the scope of his own deliberations for action. In other words, causal explanation and reason explanation are both conceptually related and practically interwoven, especially from the third person or interpreter's perspective. In a stronger, normative sense, causal explanation need not be distinguished from reason explanation for particular cases of action not only because they often coincide with each other, but also because for ultimate practical purposes, viewed from a long-term perspective, the distinction makes no real difference. The reason part, if not coincidental initially, will catch up with its causal counterpart, so to speak, or *vice versa*. More concretely, if certain deep-rooted, nontrivial causal items are initially somehow neglected by reflective or deliberative consciousness, their persistent effects will certainly be felt by the actor before long; therefore in a later formation of his judgment about reasons for a course of action, these causal items will most likely figure. In the other direction, if those initially neglected causal factors are nothing but transient, random ones which are in principle controllable or avoidable through certain methods, then the original justifiable reasons for action need no change or

adjustment but only strengthening, which will keep these causal factors in line with the reasons. In short, we need not worry about the possible occasional mismatch between reason and cause. Stable or robust causal antecedents can rationalize, viz., serve as a reason for, their consequents in the end. This is roughly the first essential intuition underlying Davidson's expression of the paradox.

- ¶29. Yet, however, there seems to be no reason why we could not have the perfect match between reason and cause at every moment of the course of action. Although it is an ideal state of mind, it is an intrinsically realizable state, given the potentiality of our free will as well as the nature of our cognitive faculties. Furthermore, such an ideal match between reason and cause is practically most desirable in the sense that any mismatch between reason and cause, even limited to a small range of mind, would in theory bring about troubles of some kind for the actor. Therefore, the ideal match can be set as an internal criterion for judging whether or not the terminal action is rational. Such a rational criterion, because of its ideal nature, makes criticism of irrationality possible whenever there is any conceivable mismatch of the relevant kind, no matter how imperceptible or negligible the mismatch might appear at the phenomenal level or the level of consciousness. Rational criticism becomes necessary when the problem of mismatch shows at the phenomenal level. This observation brings us to the second intuition underlying Davidson's expression of the paradox. It can be roughly put as follows. The minimal conception of rationality is the one that ought to fit everybody in their very possibility as free and creative beings after suspension or cancellation of all their individual contingent properties and relations; rational criticism presupposes the existence of such a universal and minimal rationality. Therefore, there is always a need of rational criticism against certain types of behaviors, which are often well-known and well-identified for certain features, precisely because they seem to have violated those principles of minimal rationality which constitute the rational backdrop for any possible understanding of human actions. As a matter of fact, one could hardly deny that in almost any human society there is such a necessity of judging some behaviors as irrational in a more or less similar sense.

- ¶30. Each of the above two intuitions or assumptions looks sound and natural on its own account. However, take these two assumptions together in a straightforward way and the apparent Davidsonian type paradox arises. I have argued that such a surface paradox can be eliminated by elucidating the following: the fundamentally normative nature of those basic principles must be understood against the background of their intriguing relationship to the relentless naturalistic tendencies of causal explanation and with the help of certain important conceptual distinctions such as that between ontological and epistemological dependence. This account, nevertheless, has led to a series of other intriguing questions, some of which, I believe, point to certain genuine paradoxes in normative human practice, especially in the sphere of normative interpretation of actions. But Davidson's original expression of the paradox of irrationality is too ambiguous to entail any of the above lines of thought, especially because it does not appeal to the ontological/epistemological distinction, or to possible interactions between the descriptive and the normative.
- ¶31. In the final analysis, the underlying paradoxes of irrationality in understanding another mind are not paradoxes in and of the agent him or herself, but rather paradoxes of the interpretative relationship between the subject (interpreter) and the object (agent) who are not merely similar in the biological sense but of the same category in a normative sense - hence such paradoxes characterize a real, *non-empty* yet *non-substantialist* concept. The notion of "normative interpretation" may sound bluntly paradoxical, as in the previously explicated sense that the very move of interpretation/description itself carries a distinctive normative function. The apparent problematics of this notion, nevertheless, reveals the presupposition of a higher-order methodological principle. That is, the rules for interpreting the intentional actions of others should incorporate expected effects of the rules for rational criticism to be applicable to those actions. Conversely, the principle would demand that any behavioral changes shaped by the (internalized) power of rational criticism be incorporated into the same category in objects of interpretation as that of "natural, spontaneous actions." As long as the "rational criticism" here is based on a thin concept of minimal rationality, the chances for such a higher-order principle to clash with reality are very slim, except for marginal cases such as strictly and obstinately incontinent actions or beliefs, whose genuine existence, nevertheless, is not taken for granted by everyone. But, on the other hand, there seems to be an inherent source of paradox, the root of which can be traced back to a certain ineradicable confrontation between the normative foundation of any principles for rational criticism and the ultimate constraints from nature which sometimes disappoint in their unexpected forms or details whatever good faith people might have about nature. A crucial point of explicating the

normative nature of human practice with rational activities, including criticisms, as its core is to see that human beings are not fixed or predetermined substantial entities, but self-binding and self-becoming potentialities whose creativity, though not unconstrained, defies any rigid, substantialist characterization. In light of this general remark, we may summarize our discussion above as a distinctive normative type of non-substantialistic approach to rational principles as well as to irrational paradoxes in the practical sphere of human existence.

END NOTES

1. Donald Davidson, [Paradoxes of Irrationality](#) in R. Wollheim and J. Hopkins (Eds.), *Philosophical Essays on Freud* Cambridge, UK: Cambridge University Press (1982), 289.
2. Donald Davidson, [How is Weakness of the Will Possible?](#) in J. Feinberg (Ed.), *Moral Concepts* London, UK: (1969), 113.
3. Donald Davidson, [Deception and Division](#) in *Essays on Actions and Events* Oxford, UK: Clarendon Press (1980).
4. Joseph Margolis, [Rationality and Weakness of Will](#), *Journal of Chinese Philosophy*, 8 1981, 9-27.
5. See Kirk Robinson, [Reason, Desire, and Weakness of Will](#), *American Philosophical Quarterly*, 28 1991, 295-296; and Robert Audi proposes, instead, a wider, holistic and non-intellectualistic conception of rational action: Robert Audi, [Weakness of Will and Rational Action](#), *Australasian Journal of Philosophy*, 68 1990, 276-281.
6. Donald Davidson, [Paradoxes of Irrationality](#) in R. Wollheim and J. Hopkins (Eds.), *Philosophical Essays on Freud* Cambridge, UK: Cambridge University Press (1982), 303.

7. Apart from pure logic rules, there are varieties of rules which reasonable people either invariably or most likely accept. For instance, Carnap and Hemple have argued for a principle which all rational people will accept though it is no part of inductive logic. It is the requirement of total evidence for inductive reasoning: give your credence to the hypothesis most strongly supported by all available relevant evidence. See Davidson, "How is Weakness of the Will Possible?" 112. The other example is the so-called Plato Principle: nobody willingly acts counter to what he knows to be best; I shall discuss this principle in the later part of this section.
8. This expression certainly allows for the possibility that the interpreter and the interpreted are one and the same person, i.e., one wants to understand oneself.
9. The "Principle of Charity" was first put forward by W. V. Quine in his account of radical translation; Donald Davidson nevertheless is its foremost proponent in interpreting both languages and actions. The basic tenet of this principle is to discourage judgments of irrationality. See Willard V. Quine, *Word and Object*, Cambridge, MA: MIT Press (1960), Ch. 2; Donald Davidson, *Inquiries into Truth and Interpretation*, Oxford, UK: Clarendon Press (1984). For a comparative discussion of principles of charity of different degrees of severity, see Paul Thagard and , Richard E. Nisbett, *Rationality and Charity*, *Philosophy of Science*, 50 1983, 250-267.
10. Donald Davidson, [Paradoxes of Irrationality](#) in R. Wollheim and J.Hopkins (Eds.), *Philosophical Essays on Freud* Cambridge, UK: Cambridge University Press (1982), 294.
11. One may observe that there are incontinent beliefs as well as incontinent actions; and the former, unlike the latter, belong to cognitive/theoretic domain only. How does the Plato Principle accommodate this? A simple answer is that even the belief against one's best evidentiary judgment is normally caused by some underpinning desires or wishes which are outside of, and not invoked as reasons in, the pure cognitive/theoretic domain. For various discussions about this, see David Pears, *Motivated Irrationality*, Oxford, UK: Oxford University Press (1984), Ch.9; Alfred Mele, *Irrationality*, Oxford, UK: Oxford University Press (1987), Ch.3; and John Heil, *Minds Divided*, *Mind*, 98 1989, 574-581.

- [12.](#) One conceivable reason is that conation or volition as functional items of mind are closer to the side of sheer causal forces and thus harder to submit to normative rules that govern propositional content and relations. This is in fact an independent topic the full discussion of which is beyond my present concern.
- [13.](#) My midway solution to Davidson's paradox here, however, distinguishes itself from a prima facie similar solution suggested by David Henderson whose approach is to accept all attributions of explicable irrationality even without a complete causal explanation of it: the latter solution does not stress the normative nature of irrationality attributions, nor does it draw on the distinction of incoherence at different levels. See David Henderson, [A Solution to Davidson's Paradox of Irrationality](#), *Erkenntnis*, 27 1987, 359-369.
- [14.](#) For the discussions of self-deception, see Jon Elster, *The Multiple Self*, Cambridge, UK: Cambridge University Press (1986).
- [15.](#) "Intrinsically irrational," in contrast to externally irrational, means that the criteria by which we judge an action irrational are not exogenous to the action itself, e.g., to see whether it is not a valid means to a certain established end, but rather endogenous, viz., to see if it in isolation violates any basic rational principles.

[Orientation](#) | [Announcements](#) | [Editorial Board](#) | [Bibliography](#) | [Call for Papers](#)
[IJFB Home Page](#) | [Submission Guidelines](#) | [IJFB Home Page](#)

ISSN 1548-6001

Professor Lik Kuen Tong, Ph.D., Editor
Professor Curtis R. Naser, Ph.D., Co-Editor
Professor Theresa Dykeman, Ph.D., Managing Editor

IIFB - Fairfield University - Fairfield, CT - 06430
Tel: (203) 254-4000 Ext. 2857, 2851 Fax: (203) 254-4074
ltkong@iifb.org -- cnaser@iifb.org -- <mailto:cnaser@iifb.org>

[About the IJFB Web Site](#)

Site Design by [Commonwealth WebWorks](#)
Database to Web Programming by [Eidos Virtual Web](#)
Powered by *ColdFusion*