Norms and Causes: Loosing the Bonds of Deontic Constraint [1]
James Swindal, Duquesne University

Abstract

Some philosophers have developed comprehensive interactive models that purport to exhibit the various normative constraints that agents need to adopt in order to achieve what otherwise would be an unattainable and unsustainable social order. Robert Brandom's semantic inferentialism purports to show how a rational construction of social coordination is enacted and maintained through specific mappings that agents make of each other's commitments (beliefs) and entitlements (justified beliefs). Strongly influenced by Brandom's account, Joseph Heath reconstructs a number of historically emergent deontic constraints that solve what are otherwise unsolvable game-theoretic problems in the maintenance of the social order. But both accounts omit a sufficient analysis of the way in which individual agents, who comprise the normative order, are effectively *addressed* by norms when they act. How does an agent, who is facing a unique interactive situation with more than one normative path to choose, make a decision? One solution, attractive to some continental thinkers, is to appeal to an innate irrational component of decision-making that lies outside of rational bounds (e.g., Nietzsche's will to power or Adorno's *das Hinzutrentende*). The model I will defend lies in an existential account of agency that occupies a middle ground between a pure naturalism (where instinct dominates) and a pure regularism, or "normativism" (where reason dominates). The existential model asserts that the given normative field within which an agent operates *conditions* the formation of the agent's intention to act but does not *determine* the effecting of an action as such — whether individual or collective. On this model, the specification of the acting or not acting on the normative intention is determined only retrospectively on the basis of what the agent actually *did* in a way that is in principle public and observable. Thus the content of the agency can be reconstructed only historically. The embodied character of the agent is what makes the action relatable to the sum of conditions that were co-determinative of the action at the time it occurred. The advantage of this view is that it does not overreach the highly limited access that we have to the inner workings of intentions to act while at the same time providing an account of agency independent of simply the agent's relation to norms.

Some social philosophers as of late have developed holistic models of socially developed and sustained normative systems. Jürgen Habermas, Robert Brandom, and Joseph Heath have developed theories which place rational choice analysis within a broader social holism.[2] Such theories, when fully developed, can actually serve to neutralize the need for strongly

---

[2] For an example of Habermas's commitment to a social holism, see his *Theorie des kommunikativen Handeln*, vol. II (Frankfurt am Main: Suhrkamp, 1981), pp. 583-593. He reconstructs the normative foundations of a critical social theory predicated on the universal structures of the lifeworld.

interventionist legal and economic steering of social coordination and thereby to support more cultural and hermeneutic forms of social integration. A problem, though, is that their reliance on forms of *deontic* constraint to solve coordination dilemmas cannot sufficiently account for the role of agent-centered purposive actions in such normative systems. Deontic constraints are determinations of consistency or inconsistency that can be applied to semantic and inferential expressions of decisions to act. For example, if I am deciding whether or not to give a contribution to a charity, I can express the proposition to do so relative to whether it would be inconsistent not to do so given other commitments I have de facto adopted by my prior actions, such as giving in the past.

In this paper I shall argue that, although deontic constraints serve to make coordinating actions possible by supplying what is underdetermined in many prior decision theories, they fail sufficiently to account for the unique contributions of individual actors to development and change within the broader normative fields in which they act. Instead, I shall maintain that a model of agent causation can both account for the originality of action and can make coherent a dynamic normative holism. Moreover, I shall suggest that such a model can actually expand the explanatory range of a normative account. My argument will proceed by reconstructing various aspects of Brandom's and Heath's models of normative holism, and then indicating how their models elide crucial causal pre-conditions for the coherence and extension of their models. Then I will present a model of agent causation that, I argue, pre-conditions the intelligibility of free action on the basis of which a dynamic normative model can even be constructed.[3]

Normative Holism

The publication of Robert Brandom's *Making it Explicit* provided significant resources for social theorists and philosophers who construct normative models of social cooperation. Brandom works from the assumption that the very "significance of being committed to a certain claim or assertible content is normative." [4] But he argues that the analysis of the interaction of beliefs and desires alone cannot provide a sufficient account of the phenomena of social action: an inclusion of deontic statuses is also needed. By including them, he constructs an expressive theory of the origin and function of norms that is meant to capture the rationality not only of basic actions, but also of beliefs, speech acts, and all forms of social cooperation. The determination of any single normative commitment can thus be situated within a wide and coherent set of such commitments. Only such a holistic theory of normative statuses, he concludes, can supplant a "gerrymandering" solution to social order that merely constructs ad hoc ways to regularize performance options rather than finding a "correct" solution.[5]

Joseph Heath's *Following the Rules* makes an even more comprehensive case for such a normative holism. Heath starts with the basic components of decision theory — beliefs and desires, on the one hand, and states and outcomes, on the other — and then demonstrates how

---

[3] Kant distinguished sharply between mathematical connections in a series, in which every condition is itself a part of the series, and a dynamical condition, in which exogenous, "purely intelligible" conditions can be allowed. See *Critique of Pure Reason*, tr. A. Wood and P. Guyer (Cambridge: Cambridge University Press, 1998), A530/ B558.

[4] Robert Brandom, *Making it Explicit* (Cambridge MA: Harvard University Press, 1994), p. 260.

[5] *Making it Explicit*, p. 208.

deontic constraints are required to make the mechanisms of instrumental (means/ends) reasoning produce rational decisions in complex social situations where agents lack either adequate information or sufficient intersubjective assurances.[6] Like Brandom, Heath refrains from abandoning instrumental or strategic models of action as such, but rather shows how they can function in game theoretic situations if tempered by certain axiomatic deontic constrains. Deontic constraints guard against otherwise inevitable collective action and regress problems that occur in instrumental reasoning. Heath's tactic is to begin with certain "folk psychology" assumptions about action (e.g., that we act on the basis of beliefs and desires), and from them to develop an expressive account of normatively guided action that fits with ordinary experience. His model exhibits the constraints but refrains from assembling them into a formal system from which predictive inferences could be made or definitive arguments to support them could be formulated.

The basic conceptual units of a normative system are *rules,* specifically the rules governing not so much beliefs or desires (as to either their origination or function), but action as such. Heath is convinced that representational models of beliefs are unable to furnish a proper account of the origin of such rules.[7] This is the basis for his rejection of *psychological imagism,* the claim that meaning is connected to representation in private mental images, and endorsement of *psychological sententialism,* the need for meaning to be expressed via predications and inferences. Heath is here consistent with a broadly constructivist view of meaning. Kant, as an example, defended a constructivism that rejected cognitive imaging in favor of a temporally linear view of perception as a continuous figuring.[8] Heath finds rules, then, as generated by social actors through their social interactions. Thus he develops a quasi-naturalistic account of the emergence of beliefs and desires within the evolution of our actual cultural practices, explicates the "master concepts" implicit in them (sympathy, virtue, and duty), and then shows how propositional language developed to assist social coordination relative to the beliefs and

---

[6] In the rational choice literature, these are generally understood to be *axioms*. "Axioms hold independently of what would happen in terms of calculating expected utility in the long run … e.g., every decision problem can be transformed into a decision problem with equally probable states in which the utility of acts is preserved." See Martin Peterson, *Introduction to Decision Theory* (Cambridge: Cambridge University Press, 2009), pp. 73-75.

[7] Like Rorty and others, Heath also takes Kant to be an image-based representationalist. For an objection to this claim, see footnote 22 below.

[8] *Following the Rules*, p. 103. So, like Dummett, he assumes that "thought is communicable without residue" (ibid.). But Heath's attribution of psychological imagism to Kant is problematic. In the Schematism section of the *First Critique*, Kant is quite clear that the bases of concepts are not momentary images, but temporalized schemata (in a linear successive time framework). He in fact strictly distinguishes schema and image. The former is a "method" of representation, a kind of macro image not reducible to specific momentary images: "Indeed it is schemata, not images of objects, which underlie our pure sensible concepts. No image could ever be adequate to the concept of a triangle in general. It would never attain that universality of the concept which renders it valid of all triangles. … The schema of the triangle can exist nowhere but in thought. It is the rule of synthesis of the imagination … The concept 'dog' signifies a rule according to which my imagination can delineate the figure of a four footed animal in a general manner, without limitation to any single determinate figure such as experience, or any possible image that I can represent *in concreto*, actually presents. This schematism of our understanding, in its application to appearances and their mere form, is an art concealed in the depths of the human soul." See *Critique of Pure Reason* A 141/B 180 - A141/B 181. In fact images themselves, which do exist, are products of the schemata, not vice versa. I am constructing the same order here: individual actions produce norms, not the reverse.

desires.[9] Like Brandom, Heath argues that the subsentential, sentential, and inferential analysis that we now can reflexively employ relative to our language use bestows us with a power to sharpen and improve coordinated outcomes of our actions by allowing us to represent the "content" of the rules, even moral ones, that guide actions.[10]

Heath knows that adjustments must be made in the theoretical understandings of both beliefs and desires in order to make this normative system work. He understands, like Brandom, that beliefs are fundamentally commitments regarding certain expectations about future states, while desires are oriented towards attaining or preventing certain outcomes. Beliefs are propositionally structured, in order to exhibit many of their properties, such as their two fold status as either *de dicto* or *de re*. He strongly refrains from endorsing a non-cognitivism about desires, rejecting Hume's belief that desires are mere "givens." In this Heath follows most contemporary Bayesian rational choice theorists. He holds that desires can be adjusted cognitively, particularly given the discounting of them that we impose as we continually face new probabilities as to whether they can be fulfilled. [11]

What we are left with, in both Brandom's and Heath's accounts, is a neo-Parsonian normative holism.[12] For Brandom, norms are applicable to acts, and hold the key to their own development and modification through the "game of giving and asking for reasons." Interlocutors keep track of the *statuses* an act reveals of an actor's commitments and entitlements via the *moves* he or she makes by verbal or nonverbal actions. Language functions to lend intelligibility to specific actions by expressing the commitments they embody and the entitlements they license. Heath gives the entire rational process a transcendental justification, inspired by Kant's transcendental deduction. Kant had confronted the skeptic not by giving arguments that would refute him, but by showing how, under the epistemic conditions to which humans have "access," certain models are transcendentally necessary in order to explain what occurs. According to Heath's justification, then, the skeptic either mistakenly thinks we have access to things we really do not or fails to see that our patterns of thought to which we do have access are constrained in certain ways. [13]

Norms and Causes

---

[9] Joseph Heath, *Following the Rules* (Oxford: Oxford University Press, 2008), pp. 121-123, 217. For a discussion of his reconstruction of a socio-biological basis for ethics, see his "Three Evolutionary Precursors to Morality," *Dialogue* 48 (2009): 717-752.

[10] *Following the Rules,* p. 265.

[11] *Following the Rules*, pp. 234-241.

[12] See, for example, Talcott Parsons, *The Social System* (London: Routledge & Kegan Paul, 1991), pp. 15-16. Parsons defines norms as "patterned interactive relationships" that make intelligible both the position of the action in the social system and the status, or role, of the actor.

[13] See *Following the Rules,* pp. 214- 215 and Heath, "The Transcendental Necessity of Morality," *Philosophy and Phenomenological Research* 67:2 (2003): pp. 385-389. But Heath does not consider the difficulties attendant upon justifying or proving what we indeed do or do not have access to. Kant, for example, assumed that we had no access to objects of intellectual intuition, only to objects of sense intuition. Such intellectual intuition would constitute knowledge "altogether different from the human." See *Critique of Pure Reason,* A278/ B334. But the assumption that there is no intellectual intuition is itself not subject to a transcendental argument.

Brandom's conceptual holism thus provides an analysis of how a normative system develops itself over time. In other words, he explains how the system can account for heterogeneous states that it renders intelligible even though the states initially form no intrinsic part of the normative integration as such.

On Brandom's view, Hegel accounted for this development by reconstructing not only the social constrains on action, but also the semantic intelligibility of them. Hegel had rejected Kant's efforts to find the source of concepts solely in *a priori* subjective activity that culminates in judgment.[14] Instead he argued that the conceptual realm emerged in tandem with the evolution of actual human social practices. The conceptual is grounded primarily in determinate negations (which Brandom calls *material incompatibilities*) that are the disjunctions of certain pairs of concepts that effectively become the impetus by which the normative system adapts to change.[15] A key incompatibility takes place at the level of the status of the social actors as such. While the master comes to realize his dependence on the slave for self-consciousness, the slave comes to realize lordship as lordship — which is not a stable reciprocal recognition. They become opposite and identical at once — and are not ostensibly reducible to an equilibrium. They become part of a set of reflexively yet historically evolved social achievements.[16] So rather than jettison the role of concepts, or rule them as incomplete for coordination, the social theorist can argue that another realm of the conceptual is demanded for rational action — a realm that extends to an emergent set of conceptual norms that supervenes on but is not determined by the empirical.

Brandom, for his part, extends Hegel's model by articulating how individual actors "keep score" of the diverse deontic statuses of those with whom they interact. He argues that one can accept Hegel's holism at the level of concepts and still critically analyze the concepts at a metalogical

---

[14] For a critique of Brandom's particular attribution of holism to Hegel, however, see Stephen Houlgate, "Phenomenology and De Re Interpretation: A Critique of Brandom's Reading of Hegel," *International Journal of Philosophical Studies* 17:1 (2009): pp. 38- 40. In contrast to Brandom's claim that Hegel understood the truth of things as being freed from their opposites, Houlgate sees Hegel as holding that things turn into their opposites such that things never are what they are. But both of their interpretations concern the holism of perception and knowledge, not of action *per se.*

[15] *Making it Explicit*, p. 160. See also Brandom's *Articulating Reasons* (Cambridge: Harvard University Press, 2001), pp. 33-34, and "Sketch of a Program for a Critical Reading of Hegel: Comparing Empirical and Logical Concepts," *Internationales Jahrbuch des Deutschen Idealimus* (2005): 140.

[16] For the reconstruction of Hegel's arguments, see Brandom, *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality* (Cambridge: Harvard University Press, 2002), pp. 178-209. But Allegra de Laurentiis argues that Brandom reads Hegel's theory of the concept too narrowly, attributing to Hegel the view that concepts operate only in the public, empirical domain of intelligibility associated with Kant's concepts of the understanding. See her "Not Hegel's Tales: Applied Concepts, Negotiated Truths, and the Reciprocity of Unequals in Conceptual Pragmatism," *Philosophy and Social Criticism* 33:1 (2007): 83-98. She claims that this mistakenly entails that the concept in Hegel refers to the sum total only of empirical concepts. Instead, she reconstructs Hegel's account to show that his domain of the concept includes non-empirical concepts (concepts of reason, e.g. freedom, spirit, God) and the concept of the subject as such, as self-actualized by its unified self-knowing and self-willing. As for the employment of concepts, while Brandom insists that concept use applies to a process of negotiation between self and other on the basis of attribution and acknowledgement of each other's commitments, she argues that for Hegel the dynamism of social interaction emerges in the *asymmetry* of the relation of master and slave that is never fully overcome. De Laurentiis puts this non-normativism quite bluntly: recognition in the *Phenomenology* is "directed to the activity and power of self consciousness in its role as 'other,' not to another individual's concept-use or rule-acknowledgment." (p. 91).

level.[17] Rational agents keep score of each other's statuses as either committed or entitled to contents that can be either inferentially or noninferentially secured.[18] These commitments or entitlements can be either *attributed* to another or *acknowledged* for oneself. An actor can, however, attribute a claim to another without acknowledging it for itself (e.g., in cases where the attributor realizes that one to whom it is attributed might not have had the same non inferential perceptual entries as the attributor). The non inferential claim of a reliable reporter, however, can authorize another to undertake it inferentially. But no knowledge is attributed to another until entitlement also is; no truth is attributed until it is acknowledged.[19] Thus Brandom's model has the effect of decentralizing epistemic authority and revealing how we negotiate rationally in a world where agents have unique and varied experiences on the basis of which to form their unique doxastic (empirical) beliefs and practical desires. Agents continually assess commitments on the assumption that other agents respond in this deontic fashion to both theoretical and practical inputs.

Brandom thus maintains that "it is the scorekeeping social practices that actually govern the use of an expression." [20] The interpreter accords cognitive authority to claims based on circumstances. The "authority" needed here is that of having an inheritable entitlement: the sort that supports successful deferrals by others (potentially including the interpreter). One cannot be said to have knowledge unless one attributes to one's own performance the significance of being a commitment involved in the giving and asking for reasons. First person deliberation is thus only an internalization of third person normative assessment.[21] Whatever self-knowledge an agent gains would be propositionally formulated and vulnerable to challenge — thus the holism is maintained.

Given this shared externalist reading of scorekeeping practice, what are the presumed conditions on the basis of which one can be recognized as a holder of normative status in the first place? One principally has to be a concept user. But if concepts are always caught up in inferential chains between at least two interlocutors, how is a conceptual chain started? Brandom at least opens up a preliminary space for the non conceptual and non inferential origin of concepts by claiming that a causal dimension of acting for reasons stems from acknowledging a practical commitment *by* acting on it.[22] Acknowledgments of practical commitments serve as "stimuli" eliciting nonlinguistic performances. Thus prior (or pure) intentions cause intentions in action. [23] As Brandom notes, "normative status is one thing, the attitudes of attributing and undertaking

---

[17] See Brandom, "Sketch of a Program for a Critical Reading of Hegel," p. 133.
[18] Reasoning is a form of inferential articulation; perception and action are non-inferential.

[19] *Making it Explicit,* pp. 201-202. Brandom makes it clear, though, that this is an expressive notion of truth.
[20] *Making it Explicit*, p. 213.
[21] *Making it Explicit,* p. 269.
[22] *Making it Explicit,* p. 261. Moreover, he refers to a parallel causal theory of perception: observable states of affairs causally elicit a perception. But the empirical inputs have to cause the dispositions in the right way.
[23] *Making it Explicit*, p. 259. Davidson sees intentions as "all things considered judgments." See his "Intending," in *Essays on Actions and Events* (Oxford: Oxford University Press, 2001), pp. 83-104. But according to Brandom, Davidson fails to show how the implicit commitment is made propositionally explicit. He thinks Davidson models action determination only on a special speech act of "shall" that makes for a promise. For Brandom, this entails that there is no sanction if an intention is not performed, thus if one changes an intention there is no need for an explanation or apology.

those states, the alteration of which is what scorekeeping consists in, is another." [24] So causation in his account makes possible the externalization and public availability of internal states so that they are available to scorekeepers. It is enlisted to distinguish between perceptual reports and actions that are not entitled on the part of the agent (merely caused) and those so entitled (caused by reasons).[25] But these causal explanations are attributed only *post facto*. [26] In perception, one is understood *to have acknowledged* a non inferential perceptual input in a perceptual belief; in practical action, one is understood to have responded to a set of acknowledged commitments by acting on them. So no *ex ante* deliberative thinking or intention formation causally plays a role in bringing about an action to which a normative status can then be subsequently attributed. Consequently, his account is not clear about how or even whether an agent *decides* to undertake these attitudes.

Heath for his part, however, drops any sustained analysis of the interaction between causes and norms altogether.[27] But by so doing, he effectively leaves deontic statuses with an even more rarified role: he does not even accept them as mediations between beliefs and desires relative to actions. Although desires can be described as intentional states, talk of intentions is at best only a part of the expressive vocabulary in the game of giving and asking for reasons. To illustrate the impotence of intentions, he refers to Gregory Kavka's Toxin Puzzle case: an interesting variation of Newcomb's problem in decision theory.[28] An eccentric billionaire wants to offer someone a million dollars merely to form the intention to take a toxin that would debilitate that person for one day only. The intention can be measured by a special brain scanner. If the intention is formed, then the person will receive the money even before the toxin is consumed, thus allowing for defection. The question is whether the intention could even be formed by the agent. Heath thinks that it is not only psychologically but also *logically* impossible to form the intention. (I shall assume, below, that is logically possible to form these intentions, but argue that they cannot be agent causes.) As he notes, "there is no fact of the matter as to whether the agent has a particular intention or not … Who has what intention is determined by deontic scorekeeping — commitments and entitlements both ascribed and acknowledged." [29] He concludes that intentions are in no way causal but are only ascribable deontic statuses derived only post facto from cases where an agent has refrained from making a typical or convention move. They are, in Sellars's

---

[24] *Making it Explicit*, p. 260.

[25] See *Making it Explicit,* pp. 260-262. He justifies this by a performative, contrary to fact, argument: if language users did not alter attitudes according to the shifting deontic statuses of their claims, there would be "no point in interpreting them as engaging in the practices specified by those proprieties of scorekeeping" (p. 260). Such a performative justification has difficulties of its own, but I shall not take up those problems here.

[26] *Making it Explicit,* p. 268.

[27] At one point in a section on weakness of will, Heath does talk about "attention management" and "self control." But these seem to be overwhelmed by the sheer rationality of re-optimizing on the basis of not exponential but hyperbolic time discounting. See *Following the Rules,* pp. 246-254.

[28] See Gregory Kavka, "The Toxin Puzzle," *Analysis* 43 (1983): 33-34, as quoted in *Following the Rules,* p. 164. See, also, Andy Egan, "Some Counterexamples to Causal Decision Theory," *Philosophical Review* 116:1 (2007)*:* 93-114 for an incisive account of problems with basing decisions on the best available (event) causal knowledge of the world we have at a given time.

[29] *Following the Rules,* p. 165.

idiom, "mongrel concepts." [30] They most certainly, for Heath, are sources neither of actions nor, even indirectly, of norms.

The problem with both Heath's and Brandom's accounts is that they do not account sufficiently for the origination of the normative system they reconstruct. A normative order cannot be a given or an end in itself, as if rationality could then be derived afterwards from it. The order cannot simply be imposed on agents without regard for the specific agent ends it would foster or hinder. So my quarrel with Brandom is that his reconstruction of the causal role of attitudes towards states and rules is incomplete, lacking an analysis of deliberative intentional aims: whether or not agents do shift attitudes remains in his system only a contingent fact.[31] With Heath, the difficulty is more problematic: individual action is evaluated relative fundamentally to expected social utility. [32] In what follows I shall explicate the form of agency implicit in the purposive actions of individual agents that in turn both presupposes and necessitates the normative systems that coordinate the purposive actions.

Non Reductive Agent Causation

Rational choice theories do not purport to describe how agents *actually* act, but rather how they *would* act if they were ideally rational. This ideal model does have some explanatory power: it has informed a number of beneficial social and political policies since its development. But even as ideal, it is incumbent upon such a model to explain *irrationality* in some measure. Most rational choice models do, in fact, attribute agent "error" to either (i) poor inferences (usually based on the agent's inadequate information), (ii) weakness of will in the agent, or (iii) factors exogenous to the agent such as the unforeseen intervention of nature to thwart its intended outcomes. The theory can then account for each source of error. It can correct a poor inference (i) by reconstructing the logical form of actions as such and then reconstruct proper inference patterns. It can correct weakness of will (ii) by showing it to be a result of poor inference regarding the combination of the agent's general aims and its specific resolves. But if lack of success still occurs when the inference patterns are correct, then rational choice theorists attribute the failure (iii) to "nature" or other exogenous factors out of the control of the agent or agents. The problem is, though, that no repair of (iii) can be proposed from within the model itself.

For Heath, outcomes in decision theory are always probabilistic given that the state or states associated with a successful outcome are "chosen" not by the agent but by nature.[33] After all, the probabilities utilized in decision theory are dependent on the chances not that the agent will or will not act on its best rational interest at the time (since it assumes the agent will act *only* on its best interests), but that certain states of affairs that influence the outcome will or will not occur.

---

[30] *Following the Rules,* pp. 164-165.

[31] At times Brandom resorts to a non causal language with regard to the efficacy of attitudes and acknowledgements: "Any effect that such elements of the score have on what performances are actually produced is indirect, *mediated* [emphasis mine] by the attitudes of those who keep score." See *Making it Explicit,* p. 260.

[32] Heath does, however, mention causation, and obliquely agent causation in his earlier, *Communicative Action and Rational Choice* (Cambridge: MIT Press, 2001)*,* pp. 53-54. See my criticism of this in "Can a Discursive Pragmatism Guarantee Objectivity? Habermas and Brandom on the Correctness of Norms," *Philosophy and Social Criticism* 33:1 (2007): 113-126.

[33] *Following the Rules,* p. 16. See also, *Communicative Action and Rational Choice*, pp. 52-59.

In such decision models, nature thus serves as an unpredictable but (usually) law-governed set of possible constraints on desired outcomes — whether they are in the form either of physical states or of states determined by the unpredictable actions of other agents. Norms are thus non-natural entities that serve as bulwarks to stabilize the expectations that agents can have of each other's actions so that coordination is possible when absolute assurance about states is not. But why is nature accorded so much power over outcomes in this scheme: a power that then can be countered only by deontic norms?

A model of agent causation, on the other hand, provides an alternative understanding of nature which is not in competition with desires for outcomes. In this view, nature — as the total set of what exists temporally (dynamically through time as states) and spatially (presently as events) — competes with neither the desire nor the outcome, but rather informs the actuality of both. Nature no longer simply effects the unpredictable states of affairs that norms function to minimize, but is itself what necessarily contributes to the action's occurrence as such. The challenge, however, is to try to formulate a view of nature in rational action that does not run into a physicalist reductionism such that nature is simply *defined* as whatever agents do.[34]

To meet this challenge, what is needed is an additional conceptual vocabulary to express the role of nature in agency. Just as an expressive theory tries to make sense of our instrumental actions, and does so given a *deontic* vocabulary on the basis of which it avoids game theoretic incoherencies (such as collective action problems or "races to the bottom"),[35] so an additional causal vocabulary is needed to rescue instrumental actions from decision-theoretic incoherencies (such as Buridan's Ass puzzles regarding the very possibility of arbitrating among rival subjective desires). The vocabulary needed to make this additional element of action explanation includes "limit concepts." A limit concept refers not to an object or state of affairs, but to a process or condition.[36] In my account here, agency is the crucial limit concept involving what conditions purposive action.

Agency is the process by which individual actors are able to take a desire (pro-attitude) for a future but presently nonexistent state of affairs (outcome) and actualize it via a given normatively bound background set of both physical and cultural conditions unique to the agent at time t. In rational choice terms, agency takes up the perspective not of "what is the probability that if I were to do X, then Y would be the case," but rather "what is the probability that if I were to do X, then Y would be the case given that I do X at t?"[37] The intelligibility of actions is determinable not just relative to natural or conventional rules that guide them, even if learned over time, but also relative to the original states of affairs agents *effect* by following the norms.

---

[34] A number of theorists attempt to do this. E.J. Lowe takes issue with the way that physicalists have wielded the causal closure argument to identify mental states with bodily states. See his *Personal Agency* (Oxford: Oxford University Press, 2008), pp. 12-14. Lowe argues instead for a non-Cartesian substance dualism that can posit an agent causation associated with mental states ("subjects of experience") not thereby reducible to physical states.

[35] For a good example of such collective action problems or "races to the bottom," see Heath's analysis in "Problems in the Theory of Ideology," in *Pluralism and the Pragmatic Turn,* ed. W. Rehg and J. Bohman (Cambridge: MIT Press, 2001), pp. 169-176.

[36] This term is attributable to Kant. See *Critique of Pure Reason,* A 255/ B 311. The noumena is an example of a limit concept.

[37] Peterson, *Introduction to Decision Theory,* p. 194. Emphasis mine.

This inverts our common conception: rather than begin with unconstrained or unsuccessful actions and subsequently develop rule based constraints on them to secure coordination, agency theory reconstructs the coordinated background factors that condition purposive but *de facto* unpredictable actions. It is not nature but the action itself that carries the element of unpredictability.

In an agency view, background norms are always already integrated both *cognitively* and *materially* in the action. Cognitively, the agent's projected purpose has to be something that is *meaningful* to the rational agent given the way the world now normatively stands *according to the agent*. An agent simply cannot act to bring about non meaningful ends because it would lack the motivation to do so. Materially, the bringing about of the outcome via the action must rely upon the physical states of the world and of the actor's body. Agents, who require bodily exertions to bring out outcomes, cannot form intentions to do actions that are impossible physically, such as to travel backwards in time or read other minds. So if there is a failure in reaching an intention, it is a function of the failed conception of the cognitive and material normative conceptualizations involved and not any kind of unprompted or unknown intervention by nature as such (such as in deviant causal chains). Agency is the achievement of acting on the basis of both limit-conceptual processes. So, this dual set of norms or rules thus *precedes* any action, not the other way around.

An example can illustrate this. Imagine an agent trying to choose between two foods to eat: one healthy and one unhealthy. On any view of action, there are at least four possible regress-instantiating instabilities a deliberating agent needs to solve in order to render a desired outcome rational to act on:

> (a) conflicts among the agent's desires for and aversions from the foods,
> (b) conflicts in the agent's attempt to find a fit between the its present desires/ aversions and past plans it had formed (such as to eat only healthy foods),
> (c) probabilities that the foods might be either inedible or unavailable (the problem of scarcity),
> (d) inabilities to secure coordination with persons who might be needed to acquire the food or to garner protection from persons who might work to render the food unavailable.

Heath's view of deontic restraint could solve each of these. (a) would be solved by stipulating that desires and beliefs are not static but fungible because of socialization and the development of the "evolutionary precursors" of duty, virtue, and sympathy. These work to limit problematic variations among desires. For example, one can begin to develop a kind of internal duty to eat healthy foods. (b) would be solved by the fact that *hyperbolic discounting* can diffuse conflicts between past plans (resolves) and present desires to change the plans in light of new desires. For example, the agent can discount the ill effects of the unhealthy food now, given that he or she has not eaten any in quite a while. (c) involves the problem of constraint by nature. Such constraint actually drives the cultural evolution forward that, in turn, guards against nature's own unpredictability. For example, sympathy and virtue evolve to enable the action coordination each needs to meet future threats of scarcity. Finally, (d) would be solved by explicating proper modes of language use and other forms of communication among agents regarding either the exchange

of basic information about the action situation or a second order analysis of the consistency of any deontic norm involved.

An agent causation view, on the other hand, could account for each of these four possible regresses in a different way. The clue for such a view is actually found in the early Davidson, who held that "the intention with which x did y," is not principally a description of an entity, state, disposition, or event, but is "syncategorematic."[38] Such a limit concept, as not itself an event but a process, does not have direct causal relations to various pre-existent states (or events) of belief or desire (a) yet cannot project an outcome without reference to them.[39] Agents are conditioned but not causally determined by their past cognitive and natural histories. Second, agents do not conceive of prior plans (b) as de-temporalized *ad hoc* constraints, but as material conditions that enable possible actions without being sufficient conditions for them. For example, if I previously gave up calorie laden desserts but am offered one as a guest in a situation that prompts my eating it, I am neither reoptimizing nor discounting the prior resolve but rather responding purposively to a present state that is nonetheless materially connected to the prior resolve to abstain. Moreover, the agent decision does not place the agent in competition with, or at the mercy of, nature regarding which dispositional state to endorse or utility to maximize (c), but rather causes a physical outcome *that becomes itself part of nature*.[40] The action of the agent is thus described, most properly, not by the mental states that preceded it, but by the physically describable outcome.[41] The action technically does not fulfill a desired outcome, which, after all, is only a mental (intentional) state; rather, it produces a previously underdetermined physical outcome that is both externally realized in the state of affairs caused by the agent and internally realized (as satisfaction or lack thereof) in the agent itself. Yet the description of the outcome is intelligible both to the agent and others only within a historical sequence of prior events that nonetheless cannot in any way predict it.

By their internal and external effects, actions become materially a part of not only the unique history of the agent but, by extension, also the agent's social and physical environments.[42] Even "failed" actions become part of this history. Consider one evening, returning home, you aim to insert your key in your front door lock and miss because the lighting overhead is dim. Though the failed outcome in this case was presumably neither psychologically intended (by your prior

---

[38] Davidson, "Actions, Reasons, and Causes," in *Essays on Actions and Events,* p. 8. I am basically endorsing this claim of the earlier Davidson, even though he later repudiated this theory in light of his development of a theory of all out intentions that need not be acted upon. See also, *Making it Explicit,* p. 256. Also, "*having* an intention" could be considered either an event or state.

[39] This is consistent with the fact that agents arguably are not even phenomenologically aware of their desires and intentions as causes. See, for example, Derk Pereboom, "Is Our Conception of Agent-Causation Coherent?" *Philosophical Topics* 32 (2004): 275-286.

[40] The account here is thus clearly at odds with that of the early Habermas. Though he spoke of a "natural history of the human species," he distinguished it from those actions that represent a "cultural break with nature" and urge towards "utopian fulfillment." See his *Knowledge and Human Interests,* tr. J. Shapiro (Boston: Beacon Press, 1971), pp. 310-312.

[41] Thus we are not referring here to the classical "interventionist" understanding of agent causation as shaping nature from without. See Lowe, *Personal Agency,* p. 6.

[42] As Theodor Adorno claims in a similar vein, one ought not narrow thought and action to their "inner historicality," but rather see them within a historicality that is far broader and includes what he terms "tradition." See his *Negative Dialectics,* tr. E.B. Ashton (New York: Continuum, 1973), p. 54.

desire to open the door) nor normatively intended (as explicitly following a "rule" of unlocking efficiently),[43] you did in fact bring about a "missing the lock with the key" result for which you uniquely are the agent cause. It becomes part of your history for future action just as much as a successful unlocking would have been. Even as a failed action, it would not have been possible without a coordinated set of physical movements and a determinant intention which remains in play even after the failure (you'll try again to unlock the lock to satisfy the intended outcome). The failed unlocking cannot even be made sense of without the intention of unlocking determining it. Moreover, presumably prior to the action, you had intended and done at least some relevant physical actions (in the example above, you presumably had a lock installed, chose to have a keyed lock, locked the house that morning, and so forth), all of which form part of the dynamic history of necessary, but never sufficient, conditions for the successful or failed action and thus also for subsequent related actions. For example, the failed action might now prompt you to install a light above the door. A normative constraint — to have building codes that require lights be placed above all outside doors, for example — could eventuate. Thus an action is not in competition with nature dynamically understood, since the new state of affairs brought about by the action – that which specifies the action -- is itself natural as a temporalized and spatialized event.[44] Thus the rational determination of action ought not to be in principle reduced to "lotteries" as they are in decision theories like Heath's: decisions subject to risk because of uncontrollable or unknowable states. [45]

A significant difficulty, of course, remains to explicate exactly how an action is described relative to its prior intentional mental states of desire and belief. Are they natural objects or not, in the sense described above? If they are not agent causes, then what exactly are they? There is no doubt that mental states are existent and describable states of affairs, and are dispositional as extended over time (even when no action emerges from them). Moreover, they are not causally inert. As states they can condition, as Davidson famously noted in his early work on action, events emitted as "onslaughts" from them.[46] They themselves, though, are not actions. Davidson made this clear also in his work on the logical form of action sentences, and repeated it in his later work.[47] But beliefs and desires are caught up in causal nexuses of their own. But their causation is not of agent, but of *transeunt*, form.[48] Transeunt, or event, causation can *post facto* be referred to as what *can* be mapped in a regular law-like succession between a prior event or state (e.g., intention) and a posterior one (e.g., action), but the laws cannot fully determine the future

---

[43] Of course, it is also trivially assumed that it is normative for an agent to be allowed to lock doors, open doors with keys, and so forth.

[44] There is an extensive literature on whether actions are events. Davidson argues that, at least semantically understood, actions are events. See his "The Logical Form of Action Sentences," in *Essays on Actions and Events*, pp. 105-149. Ryle and von Wright are among those who hold the opposite view. On the account here, actions are events as defined by their outcomes, though they are uncaused by prior events.

[45] *Introduction to Decision Theory,* p. 94.

[46] *Following the Rules*, pp. 20-24.

[47] They are "primary reasons" for action, which are essentially what Davidson later calls "all out intentions." For the later reference, see his "Problems in the Explanation of Action," in *Problems of Rationality* (Oxford: Oxford University Press, 2004), pp. 103-104.

[48] Chisholm distinguished between event, or transeunt, causation, and immanent, or agent, causation. See his "Freedom and Action," in *Freedom and. Determinism*, ed. Keith Lehrer (New York: Random House, 1966). See also Dorothy Emmet, *The Effectiveness of Causes,* (Albany: SUNY Press, 1985), pp. 76-77. A thorough formulation of agent causation is found in Richard Taylor, *Action and Purpose* (Englewood Cliffs NJ: Prentice Hall, 1966).

instances that the prior event or state nonetheless conditions. From the point of view of agency, event causes are necessary but not sufficient conditions for actions — and cannot themselves predict the events they nevertheless condition.[49] Thus actions remain free as indeterminant in this way, not in an absolute sense as pure arbitration among probable outcomes.

Of course this view could lead to the immediate objection that no deliberative and free decision effectively is made by an agent, since the decision seems merely a kind of acknowledgement of an *ipso facto* unpredictable and impersonal set of conditions of the action. But consider initially how prone we are to attribute a naturalized causation to physical objects. To heat things, for example, seems to be a natural property of hot objects. Though the hot object really is not doing anything but what is natural for it, we definitely acknowledge that it is effecting or directing something in the natural world (by either origination or modification). Yet it would be odd — and in fact irrational in Brandom's and Heath's accounts — to attribute normative states to these objects (as if it were an *entitlement* of a hot object to heat). In fact, Heath refuses to attribute deontic statuses even to non human animals: neither to parrots, who can in fact give reliable verbal reports of certain sense perceptions, nor to higher order primates, who can on some accounts accurately apply concepts to objects.[50] But why should the natural directedness of actions of non human agents not reach to same directedness in human agents? From an agency view, all actions are both immanent and material in their effects.[51] The uniqueness of human agency stems not from its connection to deontic constraints (even though these are unique to human agents), but from its power to redescribe acts linguistically as related to a naturalized history and *thus* to other agents and the constraints they causally impose by their agencies. From this framework alone can actions be described either as coordinating or, conversely, as free riding. This power of redescription of actions in no way annuls their natural, or material, status.

One could then still object that a naturalized account cannot but obviate agency. After all, all agent action will still always be compatible with a totality of physical laws. Heath himself avoids this reductive naturalism by referring to deontic constraint as a whole as "culture." He is rightly wary of the naturalism of "selfish gene" arguments, and instead analyzes altruistic cultural behaviors that undermine what otherwise would be straightforward instrumental action to protect and increase one's own offspring at all costs.[52] Yet he is also interested in examining "the sort of biological structures that must be in place in order for human cultural transmission to occur, along with consideration of how these might have arisen."[53] But even he concludes that these learning structures — such as developmental plasticity and capability for learning — actually are not biological but fundamentally cultural. Cultural constraints are needed to minimize norm-disregarding actions (particularly of free riders). But from an agency view, culture is indeed complex and normatively charged, but still is a product of and is maintained by discrete purpose actions of individual agents. As Davidson maintains, actions always have in principle a complete physical description that includes reference to what was done. But while he takes this in a

---

[49] A classic example here would be when the stillness of the lake prompts someone to take a boat out on the water. The stillness is the event cause for the agency of the person taking the boat out.

[50] *Following the Rules,* pp. 126-127.

[51] As for mental acts, they have definite physical correlates, such as brain states or outputs of the central nervous system.

[52] See *Following the Rules,* pp. 170-176; "Three Evolutionary Precursors to Morality," pp. 722-724

[53] *Following the Rules,* p. 188.

hermeneutic direction, in his stipulation that the conditioning is discovered via radical interpretation, a better alternative is to view actions, even those that bring about deontic constraints, as initially *contingent* physical events.[54] In other words every action, as contingent, is not principally determined by its pre-conditions or post-effects, even though it can subsequently be judged either to be in compliance or not with a given set of norms. Yet from the point of view of agent causation, all actions (and, ironically, the acts of free riders principally) have a measure of putative normative neutrality since the conditions under which any prior norm would apply to them are always rendered underdetermined by the originary nature of the action.[55]

What about the role of language in communication among agents about actions ([d] above)? Surely language does not exist at the level of a sheer physical input, like a sense prompting or instinctual reaction. Heath rightly sees language as a species accomplishment that aids in normative development. He supplies a theory of cultural evolution that begins, most fundamentally, with the human species' "imitation with a conformist bias" that morphs into a sequence of developments: a cultural inheritance system, rule following, the development of norms implicit in practice, the development of semantic intentionality and propositionally differentiated language, and finally the species dependence on language for planning and behavior control.[56] The problem with this kind of developmental account is not that it is irrelevant, since it is actually needed for action description, but that it suppresses the first person account that conditions the description. The alternative is to reverse the process: not to see language as an emergent property of species function, but rather as functioning at the level of the language user. It is a form of transmission from one person to another to resolve either incomplete or asymmetrical information or to determine meaning for outcome projection.[57] So rather than presume that there is an *ideal* state of optimal information that should drive all agents in their linguistic attempts to decide either on probability of outcomes or on coordination or defection, the exchange of information needs to be viewed as determined as such at the moment it is taken up by the individual agent as an *actual* condition for its determination of its own satisfaction in either coordination or defection. Still, there is no privilege to intersubjective sources of information relative to an action: a person telling someone convincingly, "you will get burned if you touch that," is not *ipso facto* a better source of information than the agent's own initial tactile sense that the object about to be touched is emitting intense heat.

It could still be objected that this agent-centered view of language represents a reductive view of human communication. The disanalogies between human communication and sense input are admittedly many. Only human communication can be intentionally deceptive or provide information that transcends the present, such as regrets about the past and warnings about the future. But Heath himself, following Dummett, holds to the radically propositional nature of

---

[54] In other words, they bring about outcomes that otherwise would not have existed, yet exist such that their (upstream) prior non-existence and (downstream) later possible return to non-existence constitutes their content as actions.

[55] And specifically in the case of free riders, their actions might in fact serve to obviate the very rationality of the norms they are purportedly violating. For example, the existence of too many free riders probably indicates the failure of the norm they free ride on to be rational.

[56] *Following the Rules,* p. 217.

[57] See *Communicative Action and Rational Choice,* pp. 65-67.

language. It is ineliminatably intentional and thus bivalent: it always, in principle, predicates something *about* something. Language thus offers no immediate access to a singular meaning. But it is precisely this bifurcated quality of language that is found in perception as well: perceptions are not immediate accessions to what is perceived, but require inferential reasoning, of some sort, to be understood. It would seem, then, that the interpretation of sensory data provides a difference of degree, not kind, from the interpretation of meaning. Thus beliefs and desires, even when linguistically expressed, condition but cannot necessitate actions ex ante even thought they are determinative for description and interpretation of actions *post facto*.[58]

Conclusion

The challenge for any normative theory of action, of course, is to classify the wide variety of human actions which would have to be subsumed under it: acts of perception, acts involving interaction with physical objects, acts of appreciation of (or distaste for) perceptual experiences, risk taking acts, mental acts, and explicit acts of intersubjective coordination (e.g., political actions) — just to name a few. The account sketched here is meant not to obviate the need for normative constraints on action, but to enclose the constraints in a broader description. All actions cause natural effects and as such are, in the first instance, able to be singled out (thus *ipso facto* have a measure of contingency) and thus are extra-normative.[59] Normative systems and the deontic constraints within them are thus second order prescriptions cognitively developed precisely to allow for agents to reach agent-determined outcomes. The normative restraints cannot be ends in themselves. Culture is a repository of deontic constraint, a kind of "second nature" that, in Heath's account, fills in coordinative gaps. But in the account here, if we no longer take "first nature" to be considered a set of capricious restraints on outcomes, it becomes superfluous to give a second (deontically developed) nature a primary status. Thus an agent causation view solves the vaunted Parsonian "problem of order" by specifying exactly how nature, viewed dynamically, stands in no arbitrary relation relative to free purposive action. What should be clear is that no game theoretic coordination can resolve the problems stemming from complex situations in which agents view nature (and *ipso facto* other agents) primarily as a

---

[58] But another similar objection could be lodged against this account. One could argue that an interpretation requires a level of *triangulation* that requires another language-using agent with whom sources of perceptual chains of events can be located. For one relevant view of triangulation, see Davidson, "The Emergence of Thought," in*, Subjective, Intersubjective, Objective* (Oxford: Oxford University Press, 2001), p. 128. Only this would seem to give closure to what otherwise would produce problematic regresses concerning origins of stimuli. But an agent view would find the origin of the stimulation a rather remote problem. It is rather the agent's *taking of* the stimuli that is determinative, not the grasping of an account of its origin. But when origins are important, it is quite possible to triangulate without a second human agent: even an inanimate object, such as a measuring device, can serve to triangulate one's determination of a distal object's stimuli. Thus I am sympathetic with Dretske's view of triangulation primarily as a causative one, in contrast to Brandom who, like Davidson, sees it as necessarily involving at least two language users. See Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: Cambridge University Press, 1981). For Brandom's account, see *Making it Explicit,* p. 430.

[59] Some philosophers of action raise problems that involve failed actions in persons with bodily paralysis. Jennifer Hornsby, for example, argues that these situations serve to indicate that actions are not bodily movements, but tryings: and these tryings have no direct causal relation on any bodily movement. See her *Actions,* (London: Routledge, 1980), pp. 39-45. But the account developed here would reduce the resolve in these paralysis cases to mental intentions, but not actions as such.

potential obstacle to their securing of outcomes. Sub-optimal outcomes should be explained, rather, not as failures of a probability calculus but as derived from insufficiently naturalized desires and beliefs.

Some measure of spontaneity in an action is what prompts us to attribute agency to an agent. The somewhat cursory outline of non reductive agent causation that has been presented, though, would also still have to account for those internal conflicts in agents indeed properly described by psychological predicates. This is usually the point of the analysis of weakness of will. The account here attempts only to reconceptualize such conflicts: an inner conflict occurs not between the intentional states competing to prevail and thus determine an action, but as a permanent tension arising from the concise spatial and temporal situation of each agent action relative to a history, both past and anticipated, of action outcomes.[60] What remains primarily determinative of an individual action is thus not how it was desired or believed by the agent, but that it, in a constitutive sense, may or may not have occurred at all in the process of natural development. Thus an action's occurrence is better explained not as an achievement of an equilibrium of motivation, projected outcome, information about the physical world, or cultural background, but as a contingent and spontaneous actuation of a new state of affairs in both a cognitive and material ambient.[61] Such a theory thus could serve to form a compatibilist theory of freedom of the will.

Contact details: swindalj@duq.edu

---

[60] This assumes the somewhat problematic claim that every state of affairs is begun by a singular event.

[61] Kant foreshadowed this when he indicated the existential and conceptual orders are distinct, save the fact that the existential entails that perception, restricted by empirical laws, in principle precedes all concept formation. See *Critique of Pure Reason,* A225/B272.