I. Introduction

In 1906, Mark Twain published a small, misanthropic essay entitled What is Man? It took the form of a dialogue between two men: one old, the other young. It is characterized by a rather striking thesis: "Man originates nothing. All his thoughts, all his impulses, come from the outside. . . . None but the gods have ever had a thought which did not come from the outside" (10–11). The rhetorical strategy of the dialogue is to show, contrary to each and every intuition one has on the matter, that the human being is a determined, yet malleable, beast.

From the cradle to the grave, during all his waking hours, the human being is under training. In the very first rank of his trainers stands association. It is his human environment which influences his mind and his feelings, furnishes him his ideals, and sets him on his road and keeps him on it. . . . He is a chameleon; by the law of his nature he takes the color of his place of resort. The influences about him create his preferences, his aversions, his politics, his tastes, his morals, his religion. He creates none of these things. He thinks he does, but that is because he has not examined the matter (57).

The dialogue continues for days. The younger man, no doubt still too infatuated by the wealth of potential goodness in his fellow man, is moved time and again to propose counter-examples to the older man’s well-worn wisdom. The older man, in good dialectical fashion, re-describes the younger man’s examples in terms of his central theoretical commitments: that no man originates anything, that each agent is fundamentally committed to the satisfaction of his own innermost wants, and that training can shape these wants.

Old Man: Adam probably had a good head, but it was no use to him until it was filled up from the outside. . . . A man’s brain is so constructed that it can originate nothing whatever. It can only use material obtained outside. It is merely a machine; and it works automatically, not by will power. It has no command over itself, its owner has no command over it.

Young man: Well, never mind Adam: but certainly Shakespeare’s creations—

Old man: No, you mean Shakespeare’s imitations. Shakespeare created nothing. . . . Shakespeare could not create. He was a machine, and machines do not create. (11)
The old man’s view is a position in philosophical anthropology. He provides a theory of agency which makes sense of the actions of particular persons in light of a view concerning the relation between selves, environment, and activity. As a theory, there is much which strikes the reader as troublesome. The philosophically minded have long suspected that any theory which could explain everything—even its own counterexamples—was explaining too much; that any theory which, in virtue of its explanatory power, could not have counterexamples, was in principle problematic. This certainly seems to be the case in the old man’s thesis that the human mind is incapable of originating anything: ideas, sonnets, blueprints, or whatever. Any counterexample can be redescribed in terms of what appears to be a rule: human beings are machines, and machines do not create.

Now imagine for a moment that the great theses about human agency which litter the historical landscape all take this form: they are recommendations for understanding the significance of human action, thrust upon human agents by all-too-witting authors, aiming to provide a new way of understanding human activity: Nietzsche’s will to power, Marx’s creative capacities, Freud’s id and superego, Twain’s inner Master and self as site of determined forces. These theorists have projects resembling therapy—getting one to see the significance of what one does and who one is under new, and often disturbing, rubrics.

The theoretical possibility which lurks at the heart of this paper is that all theories of human agency might take this form—might be rules of description for understanding the significance of what humans do, rules capable of altering the very significance they attempt to capture. If this is right, there is reason to re-think the terms in which theories of human agency are discussed.

The examination which follows is meant to show why one might reconsider the nature of theoretical accounts of agency. To this end, an analytic vocabulary for discussing types of statements will be developed (section II) which will allow a specification of some of the available alternatives in considering theoretical claims. These distinctions will then be applied to a particular debate in psychology: the situationist/personality theory debate (section III). Finally, an argument is provided for thinking that theories of human nature and agency are best seen as what one might baptize “recommendations.”

II. Types of statements

Three types of statements are distinguished in what follows. These distinctions are bold and abrupt. No apologies are offered for the stipulative character of many of these remarks: the aim of this article is to provide a way of talking about certain distinctions which proves useful for re-conceptualizing approaches to various sorts of theories. These distinctions are inspired by scattered remarks in the Wittgensteinian corpus, though Wittgenstein will not be invoked to justify the distinctions in question.

To avoid confusion, it should be remarked that the distinctions which follow are not syntactical ones: any remark can be given the logical form used to distinguish the categories below, depending on the use to which one puts a particular statement. One should not suppose that a particular syntactical structure will dictate the category to which a statement belongs. Rather, the below distinctions are distinctions of use. There are several criteria by which one might distinguish these uses. While these categories are not exhaustive,
they are instructive for deciding, in a particular context, the use to which a statement is put.

The three uses of statements to be considered, and which should not be construed as complete, can be baptized “assertion,” “clarification,” and “recommendation.” To begin distinguishing these uses, consider a (highly diverse) set of sentences:

a) “The morning star is the evening star.”
b) “A sentence has a subject and a verb.”
c) “A fetus is not a human being.”

As is well known, Frege spent much time considering (a). The result, ultimately, was the sense/reference distinction.1 The two terms involved in the identity statement “The morning star is the evening star” have different senses, but share a reference.2 It is the presence of different senses that allows (a) to assert something about the world (the claim is not analytic, as in “The morning star is the morning star”). The result of (a) is epistemic gain: the statement reveals something about the world.

According to the traditional wisdom, in order for a statement to be called “true,” it must be possible to formulate an idea of what it would mean for that statement to be false. This is a necessary, but not a sufficient condition for regarding (a) as both assertoric and true. This possibility in regard to (a) is easy enough: it could be the case that the object picked out by “The morning star” is not the same object picked out by “The evening star.” If this were the case, there would be a difference in reference, and the claim that “The morning star is the evening star,” as an assertion about the identity of the references in question, would be false. If one could not imagine what it would be like for (a) to be false, one would be hard-pressed to claim that the statement was, in fact, assertoric. Saying something about the empirical world, as opposed to saying something about, say, the grammar of a sentence, presupposes that what one says could fail to map onto the world—and thus presupposes that what one says could be false.

It is precisely for this reason that an analytic claim (such as “The morning star is the morning star,” or “All bachelors are unmarried men”) cannot be an empirical claim: no empirical considerations can falsify the claim in question. To show that the statement “All bachelors are unmarried men” is false, one would not look for a bachelor who was not, in fact, unmarried. One would point to discrepancies in the meanings of the two constitutive elements in the identity-statement, or something else of the same Quinean ilk.

This is the first feature of what one might heuristically baptize an “assertion”: a statement which is used to assert something is empirical, and for that very reason, can be either true or false given empirical considerations. This feature already allows one to do something of which the positivists proved so fond: one can distinguish significant assertions from nonsensical ones. A statement which purports to assert something, but which cannot admit of either empirical evidence or empirical falsification, fails to be a significant assertion:3 it reveals nothing about the world.4 A statement which does find support in the empirical counts as “knowledge.”

The point of this excursus into an old analytic wisdom about assertions is contrastive—it is not to claim that all knowledge should be seen under the rubric of assertions—a temptation to which Wittgenstein fell, both early and late.5 Rather, this excursus allows one to demarcate a specific way of using a statement—one which can be articulated as having the form “x is the case.”

What is interesting about this use of a statement is that it constitutes a relatively small part of linguistic activity—both on
the street as well as in the academy. Philosopher and lay person alike mistake statements of all sorts for statements of the form "x is the case," failing to notice certain crucial features of said statements which point in another direction—one which looks much more like sentences (b) and (c) above.

Both sentences, (b) and (c), have the appearance of assertions. They could both be forced into the form "x is the case" by filling the variable in the claim "it is the case that x." But one should not be misled. Neither (b) nor (c) can count as an assertion, because neither sentence admits of empirical considerations. To put bluntly what will be spelled out in the remainder of this section: (b) is a constitutive rule of grammar, and (c) is a recommendation for adopting a constitutive rule for what it means to be a human being.

In The Construction of Social Reality, Searle claims that a constitutive rule has the form "x counts as y in context c." The x term, in this form, is meant to pick out either 1) some range of physical behavior (e.g., moving a leather ball from one end of a field to another, moving a wooden piece along a diagonal line) or 2) some physical object (e.g., a green piece of paper, a stack of paper upon which one finds a word banquet). The y term, in this form, picks out the significance the x term has, given a shared understanding of that which the x term picks out (e.g., a football game, a move in chess, money, a submitted article). The c term acts as a specification of the situation in which the "counting as" relation will hold: a stack of paper will count as a submission to a journal only in specific circumstances; a piece of green paper will count as money only when it has been made in the correct way, by the correct agency, and so on.

A constitutive rule makes possible a type of behavior, or a type of object. Some range of activity will count as a football game only because there is something it means for an x to be a football game. X can count as y only because of a shared understanding (an intersubjective, non-cognitive agreement) about what constitutes y.

The Wittgensteinian corpus is full of examples of constitutive rules: "Red is a color," "There are physical objects," "A language is public." These rules reflect what will count, given intersubjective agreement, as using color terms in the way they are used, sharing a world-picture, and having a language as understood by human beings. While this is a (largely) persuasive reading of Wittgenstein, one needn't participate in this conversation in order to delimit the category baptized "clarification." A statement is used to clarify something (e.g. a game, a practice) when it articulates what counts as playing a game or participating in a practice. Sentences of this type are non-empirical. They do not assert something about the world, and are not falsifiable by empirical means. Some examples are in order.

"Moving this piece of wood along a diagonal (x) counts as moving a bishop (y)." It is unclear what it would mean for this statement to be false. The statement captures an understanding of what it means to play a game of chess which involves an understanding of how each of the pieces moves in the game. The game is constituted by this understanding—without such an understanding of the range of behavior in question, there would be no game of chess at all. And what this means is precisely that a rule, when it really is a rule, is neither true nor false: it is an articulation of what agents take to constitute a type of activity, or type of object, which is made possible by a shared understanding of the activity or object in question.

But if a constitutive rule is not falsifiable, how can a statement which articulates
a rule be significant? Moreover, if a rule is neither true nor false, what is its epistemic status? These scripted questions have answers: a rule is meant to tell an interlocutor what counts as y—what is involved in participating in some y, or calling something "y." It thus clarifies an activity or an object by spelling out the "counting as" relation. Claiming that these clarifications are neither true nor false, of course, is a bit misleading. It is true, after all, that "A sentence has a subject and a verb." The point here is that the claim to truth in (b) is not identical with the claim to truth in (a): (b) is true only given intersubjective agreement, while (a) is purportedly true regardless of whether or not such agreement exists. It is this distinction which Searle usefully captures in the claim that a rule is both ontologically subjective and epistemologically objective. The sentence "Hitting a baseball out of the park counts as a home run" is true—objectively so—but it is true only because there is something it means to hit a home run. The existence of a "home run" is thus ontologically subjective (it depends on human practice), but epistemologically objective (there is something it means to hit a home run).

A statement of grammar, such as (b) above, has the form of a constitutive rule. It clarifies what it means for an arrangement of symbols to be a grammatical sentence. This is objective, but there could be no fact of the matter without intersubjective agreement about the "counting as" relation. It is true, then, but not in the sense that an assertion is true: empirical considerations will not decide whether or not a sentence has a subject and a verb.8

The third type of statement is, in one sense, a continuation of the second.9 One can baptize this category of use "recommendation," and take as one (preliminary) example (c) above. A recommendation has the form "x should count as y." It will prove helpful to distinguish this from two sibling notions.

A recommendation is not a regulative rule. A claim about how one should regulate one's behavior is not a claim about what should count as a certain object or range of behavior. When someone says "Drive on the right side of the road," she is offering guidance for engaging in a certain activity. Driving on the right side of the road does not constitute the act of driving—one can do this in fields, on racetracks, in the mud, or any other number of places. Moreover, the speaker is not proclaiming what should count as driving. She is simply providing a rule which might regulate the way her interlocutor engages in the activity in question, not one which constitutes the activity, nor one which she thinks should constitute the activity.10

Nor is a recommendation a run-of-the-mill hypothetical imperative. When someone says "One should brush one's teeth if one wants to avoid cavities," no attempt is made to change what is regarded as "brushing one's teeth," nor claiming that whatever "counting as" relation might be present in teeth-brushing might be modified in a certain way. A hypothetical imperative is not a constitutive rule, nor is it a recommendation for one.

So what is a recommendation? There are cases in which there is ontological subjectivity, but not epistemological objectivity in any significant sense. These are often cases in which one has met the limit of intersubjective agreement about constitutive relationships—where questions emerge about what will count as something else—where the domain of an x is in question. The circumstances in which these cases emerge are often of the most banal sort, where what is sought is a way to proceed through unclear cases, and nothing much rides on the decision (e.g., should vehicles x and y count as trucks? Should a banana
tree count as an herb, given certain biological features?). In other cases, the political stakes are high, and the implications of a decision staggering (e.g. should this act count as treason, should a fetus count as a human being?).

A recommendation attempts to further define the constitutive rules of a practice—be it football or classifications generally. These statements emerge when one encounters those things which do not easily fit into the mold, and aim to cover the new case by either 1) specifying the (indeterminate) rules already in place or 2) outright replacing those rules.

(C) above is one such recommendation. No one on either side of the abortion debate would deny that one should refrain from killing human beings whenever possible. There are few, if any, moral monsters among the participants in this conversation. Emotions run deep precisely because of a shared intuition about the value of human life, and an equally deep disagreement about what should count as a human life. What is contentious about viewing this as a debate between two potential constitutive rules is that it supposes that empirical considerations will not decide the matter—and, ipso facto, that what it means to be human is not an empirical question.

Defending this view could only properly be done in a book-length discussion, if at all. But it is precisely around this question that the abortion wars rage: should being biologically human be the deciding factor in what counts as a human being, or should it be a “soul,” a mental life, the potential for a mental life, or some other thing? All of these claims can be regarded as recommendations for specifying the x term in “x counts as a human being.” The participants in this debate are not concerned with what biology has to say about human existence, but about what should count as a human life—even if what should count is what biology asserts.

Dealing with such a politically charged example perhaps obscures the point here at issue. This is a worthwhile risk, as the example seems to be a perfect instance of philosophical confusion. A asserts that “A fetus is a human being.” This statement looks like a claim of the form “x is the case,” and not of the form “x should count as a human being.” A attempts to adduce empirical evidence in support of his claim (e.g. biology asserts that a human being has this genetic structure, and a fetus has this genetic structure; look and see). B denies this, say, by pointing out that the empirical evidence is misleading (a lung contains the complete human genetic code, but there are no qualms about removing it if one needs to). A debate about recommended rules—about whether or not one should expand what all agree will count as a human to include the fetus—is converted into quibbling about empirical data.

The same sort of analysis as the one above can be applied to numerous cases in applied ethics: euthanasia, questions about race, gender, or even sex. There is no need to make matters more difficult by dving into these divides. The aim here is to articulate a distinction.

The features of a recommendation which should be highlighted are the following: 1) a recommendation is potentially epistemologically objective. What gets recommended as counting as, say, a touchdown in cases where there is uncertainty about how to proceed can quickly be made into a constitutive rule. 2) A statement used to recommend a constitutive rule is almost always significant: it must be sincerely uttered, and it must be possible. 3) what decides between recommendations, unlike what decides between two assertions (e.g., p and ~p) will not be empirical. Finally, 4) What distinguishes (b) from (c)—a
clarificatory statement from one which recommends—is the degree of intersubjective agreement which attaches to the “counting as” relation.

The distinctions made here are summarized in Table A, above. These distinctions will be explored in more detail in what follows by applying them to a particular example: the debate between personality psychology and situationist analysis.

III. THE FLUIDITY OF CATEGORIES

There is, in principle, no necessary distinction between a clarification and a recommendation. Clarifications involve implicit recommendations which stem from their ontological subjectivity. When one says “x counts as y,” one presupposes a shared understanding (the y term) which allows one to understand a set of behaviors (the x term). The articulation of a constitutive rule of this form endorses that rule: by saying “this is the way it is done,” one is also saying “this is the way it should be done.”

The difference between the two categories, then, seems to lie in the phenomenological force of the respective normative claims. Recommendations are answers to questions where the status of some x is in question. Examples abound: should the fetus count as a human being? Should quantum mechanics count as explanatory? Should a ballot that has not been clearly punctured count as a vote? Should rhetorical questions count as assertions?

The category of recommendation is useful precisely when the category of clarification seems at an end—when the intersubjective agreement about the “counting as” relationship is unclear. Constitutive rules are recommended precisely when cases are encountered which do not clearly fit the mold—which may or may not fall under the rule in question.

It seems that many substantive disagreements can be cashed out in terms of disagreements about the scope and limits of certain constitutive rules. One can see these disagreements as paraphrasable in terms of different recommendations for how to proceed when encountering new cases.
Consider, for example, the disagreement between personality theorists, on the one hand, and those oriented in situationist psychology, on the other. The old wisdom about explaining the actions of an agent would have one believe that causes for action are describable in terms of specific character traits—that actions are trait-relative, as it were. The situationist challenge amounts to the claim that situational factors are more explanatory than an explication of personality structure: an agent's reason for administering voltage to a confederate has more to do with the situation in which the agent finds himself than it does with the character traits which constitute his personality.\textsuperscript{12} Now, is there simply a fact of the matter here about what will be more explanatory—and hence about the way this region of psychology should proceed—or are the warring parties in the midst of a disagreement about what to count as explaining a human action?

On the one hand, it seems that situationist psychology has made it virtually impossible for the older models of the correlations between character and action to carry the day: canonical accounts from Aristotle to Allport do not adequately meet the demands placed on an explication of human action by overriding situational factors—factors that have frequently, experimentally speaking, made the possession of specific character traits look obsolete. Of course, very few would deny that character traits have some bearing on responses to different social environments, but many would deny that these traits constitute a difference that makes a difference. On the other hand, personality theory has managed to yield models with substantial success, and situationists have again and again returned to specific empirical data which suggests that the very notion of personality traits is a problematic one: it cannot adequately explain the specific actions of agents in specific situations—situations which seem to make trait talk obsolete.

Part of the problem in the grand psychological divide between situationists and trait theorists can be cashed out in terms of different explanatory paradigms. The situationist thinks that situational factors are the ultimate unit of analysis when attempting to explicate various human actions: they take it as a rule for talking about human action. The trait theorist finds an overarching framework of character-specific traits to yield the most fruitful results when attempting to understand why an agent does what she does. The situationist thinks, pace this traditional wisdom, that one should modify what one takes to be significant: one should look to external situational stimuli in order to understand the actions of an agent.

There does not seem to be any convincing way to decide between these two models of explanation. While the trait theorist certainly has intuition on her side, intuitions which seem to stand in defiance of empirical data are hardly intellectual allies. Further, although situationist psychology has done some intellectual damage to trait theory, it has hardly made such theory obsolete: although perhaps inelegant, it is still possible to explain situationist data in terms of antecedent agent-specific behavioral dispositions. Here, one is confronted with the (perhaps not so) peculiar situation of deciding between two quite compelling, and yet antithetical, models of how one might best explicate human activity.

But how much is to be made of the disagreement?

Some the most ardent advocates of the situationist view have, in some respects, conceded the game to trait theorists. Likewise, psychologists engaged in personality theory do not claim that external situational stimuli play no explanatory
role in explicating and predicting human behavior. Quite the contrary, situational factors can readily alter those character traits one might expect to be manifest in specific situations. Take the following two cases, for example.

Ross and Nisbett, among the most famous contemporary situationists, have recently weakened what was once a die-hard anti-personality account of human behavior to accommodate an explication of individual differences in terms of character. Whereas lay psychology was once viewed as (mostly) mere cognitive illusion on the part of active agents, Ross and Nisbett now acknowledge that “the predictability of everyday life is, for the most part, real” (7). While earlier writings had suggested that the systematicity perceived in human behavior was the result of certain false conceptions of the role of character in everyday activity, recent writings suggest that the situationist challenge is most (and perhaps only) forceful when particular agents are immersed in quite new circumstances—when they “take on new and different roles or responsibilities, encounter new cultures, analyze newly arisen social problems, or contemplate novel social interventions to address such problems” (8). In these new contexts, lay psychology breaks down and situational emphasis carries the day.

But notice that this makes the personality/situation debate seem a bit artificial. The thrust of this debate centers around whether or not human behavior can best be explained and/or predicted by reference to specific personality traits. In an arsenal of experimental work, the situationist movement has attempted to undermine the rather common-sensical view that personality traits are a causal antecedent to action. But in each case, the situation presented to the individual agent was an extremely novel one—one in which the responses a personality trait might predict would be unknown due to the novelty of the situation. If Ross and Nisbett are correct to allow for predictability in terms of lay psychology, it is certainly plausible to construe the importance of situational stimuli as the result of the novelty of the specific situation, if only because what will count as a personality trait exerting a causal influence in a novel situation is unclear. But very few personality theorists would deny such a thing. Cheek and Hogan, for example, maintain that “clearly, situations influence social role performances, but mature adults also have coherent styles of life grounded in relatively enduring values, goals, and interactional strategies formed during personality development” (256).

The difference here seems to be one of emphasis. If one is to predict human behavior in novel situations, it is likely that such predictions will gain more from noticing situational variations than from articulating hard-won personality traits. In a novel situation, one doesn’t know what to count as an action stemming from someone’s character, simply in virtue of the fact that the situation is a novel one. On the other hand, if the prediction in question is about behavior in familiar and common contexts, there is probably little need to concentrate solely on situational stimuli, as Nisbett once suggested (1980), as there is a relatively stable understanding of what sorts of actions will stem from what sorts of traits.

The problem here is precisely the one seen above: in cases which are not obviously character-driven, there are two divergent recommendations for what will count as explaining, causally, human activity. The personality theorist regards the notion of an action stemming from a trait (or set of traits) as that which counts as explanatory; the situationist does not. The personality theorist wants to offer explanations of
actions in novel settings in terms of situational stimuli affecting personality structure; the situationist finds the postulate of personality in such settings superfluous. The reading of the empirical data, then, seems to be the by-product of the recommendations made in these two respective camps. The personality theorist regards the possession of a more or less coherent, more or less static,\textsuperscript{13} set of traits and dispositions as constitutive of human agents, and hence as constitutive of explaining the actions of such agents. The situationist regards the constant influx of external stimulation and responses to it, as well as the agency which is the site of this melee, as constitutive of human engagement in the world, and hence as what will count as explanatory of it.

Two recommended rules for what will constitute human action and that which explains it, and no clear solution in sight. But this debate can be complicated in an even more pervasive way. Cheek and Hogan (along with a plethora of others) maintain that interactional strategies are the product of personality development. These interactional strategies undoubtedly involve phenomenology. When one encounters a specific interaction situation, the strategies one employs (according to the personality theorist) will be the product of the significance the situation has for the agent in question, and the significance things have for this agent, in turn, will be circumscribed by personality traits. Personality traits are predictive, not simply because they exert a causal influence on what one does, but also because they exert power over what will be regarded as relevant in a specific instance.

Leaving the term "personality" out of the picture, one finds a similar view in some situationist social psychology. Even heroes in the tradition of situationist social psychology have acknowledged the importance of phenomenology. Indeed, Ross and Nisbett emphasize the importance of phenomenology by including phenomenological considerations as a basic tenet of social psychological inquiry (Ross and Nisbett, 1991). The Principle of Construal acts as a cornerstone for social psychology insofar as this principle acknowledges that "the impact of any "objective" stimulus situation depends upon the personal and subjective meaning that the actor attaches to that situation" (11).\textsuperscript{14} A situationist analysis of human behavior, then, cannot rely on a traditional stimulus-response model so common in simplified behavioristic analysis; situational stimuli are always subject to the interpretations of the human beings who encounter them.

But in what sense is the Principle of Construal not an acknowledgment of those personality conditions (which are undoubtedly culturally informed) leading to specific interpretations of given situations? It is here that differing terminology exacerbates a caricatured debate. Consider the following passage from The Person and the Situation:

individuals may behave in consistent ways that distinguish them from their peers not because of their enduring predispositions to be friendly, dependent, aggressive, or the like, but rather because they are pursuing consistent goals, using consistent strategies, in light of consistent ways of interpreting their social world (20).

The distinction here is between consistent goals/strategies/modes of interpretation and the possession of character traits. But notice that one need not make such a distinction. Indeed, Cheek and Hogan seem to claim that goals, strategies, and modes of interpretation are explainable only in terms of personality traits.

What the above case presents is a disagreement about the best way to carry on a social psychology—a disagreement not only about what should count as the basic explanatory methodology (does one take
the trait as trump, or the set of situational factors?), but also about what will count as a "personality trait" at all. What is at issue is not only what will count as explaining the data, it is also about what the data in fact is.

If this characterization of the debate is adequate, there are not two theoretical statements of the form ‘p’ and ‘¬p.’ Rather, one finds statements of the form “p (should) count as explaining human action,” and “p does not count as explaining human action.” As was hopefully made clear above, this disagreement sometimes amounts to the (more specific) claim that “p (should) count as a personality trait” and “p should not count as a personality trait” (this is the significance of the principle of construal—without the second claim, Ross and Nisbett become personality theorists). The debate does not center around the oodles of experimental data. Rather, it centers around what constitutive rules one should take as fundamental, and no experiment will decide this question.

It should be emphasized that these recommendations for rules are not to be sharply distinguished from clarifications of rules in other contexts. One can make a distinction here only because the issue has not been decided in cases where recommendations flourish. For a recommendation to emerge, there must be a question about how to regard some x: does it count as some (prior) y, or does it not? One proceeds by attempting to determine an answer to this question. In psychology as much as in a game of football, when events occur which are not straightforwardly categorizable in terms of extant constitutive rules, one makes recommendations concerning how to regard the event: A player spikes a football a split second before making a touchdown. No one could have stopped him. The question: does this count as a touchdown? It is implausible to regard this instance as different in kind (though it will certainly be different in detail) from the following one: subjects administer shock to a confederate who pleads with them to stop. The question: does this count as acting from personality traits, or responding to situational pressures (the presence of authority)? Are these considered mutually exclusive?

Because answers to these questions can further define a constitutive rule, a recommendation can quickly become a constitutive rule. “Spiking a football into the end zone, before entering the end zone, should not count as a touchdown,” as a recommendation for dealing with a specific event, can quickly morph into a constitutive rule where there is no need to articulate the strongly normative “should.”

Likewise, a constitutive rule can be problematized by recommendations. Indeed, much work done on questions of race, sex, and gender from the social constructionist sector takes precisely this tack: they problematize a specific “counting as” relation (e.g. being this color counts as being of this race, fulfilling these roles counts as being of this gender, having these gametes counts as being this sex) by transforming a rule into a recommendation—by showing that what one has taken to be constitutive of some x (morphology being constitutive of race) is optional and, often, to be avoided.

The categories of clarification and recommendation, then, are fluid. Distinguishing the two allows one to see a different normative force attaching to different types of statements, but it does not allow one to say definitively which statements will always be recommendations, and which clarifications.
IV. Theories of Human Agency as Recommendations: A Preliminary Argument

Theories of agency have often been regarded as falling in the first category adumbrated above: when persons discuss agency, they assert facts about the world of the form “x is the case.” The (therapeutic) proposal of this article is to cease and desist from this practice. Instead of viewing theories of human agency as assertoric, one can fruitfully view them as either clarificatory of how agents in fact proceed or as recommendations for how agents ought to proceed. Deciding between these two alternatives will often be a rather tricky matter, and one need not insist that a particular theory falls squarely into one of the above categories: most cases, it seems plausible to concede, will be borderline ones.

This fact notwithstanding, the above distinctions make clear the ways in which a theory of human agency will differ from theories about, say, molecular biology. Typically, a theory is regarded as some set of propositions which explain, or attempt to explain, some datum, or data. These propositions purport objectivity of some sort or other: theories are meant to be spectatorial accounts of some object or range of behavior, and it is this spectatorial quality which makes meaningful claims to objectivity from theoretical perspectives.  

A theory of human agency will thus be a spectatorial account of human agents—one which aims to articulate regularities in the actions of those agents observed which would be generalizable across the domain of human actions.

Now, consider two hypotheses about situated human agency which do not seem atrociously implausible.

1) Human agency has social conditions.
2) Certain theories of human agency display a “looping effect.”

These will each be considered in turn.

The first hypothesis might be called “the developmental thesis”: in order to become effective agents, certain social requirements must be met. These social requirements can be distinguished into (at least) two types.

Type A: Basic Sustenance Requirements. This category is not meant here to pick out the range of non-social requirements which are required for the development of agency (e.g., nutritional needs, shelter, clothing, etc.). Rather, the category designates a set of (relatively straightforward) social requirements for sustenance. These range from interaction with other competent language-users to physical contact. Failure to satisfy Type A requirements can lead to failure to develop agency, or, more commonly, to damaged agency.

Type B: Phenomenological Requirements. In addition to basic sustenance needs, physical and social, agents have “phenomenological” ones. The range of practices which a theory of agency hopes to cover, and in which human agents engage, are constituted by an intersubjective understanding of the meaning of the activities within the practice in question.

An agreement about how one scores a touchdown, for example, is essential to playing football.

Now, as should be obvious, there are significant asymmetries between the two types of social conditions. Without satisfying requirements of Type A, an agent never has the opportunity to satisfy requirements of Type B: if one does not possess linguistic competence, one could not begin to understand what will count as voting, christening a boat, or engaging in the stock exchange. On the other hand, without understanding how to engage in the practices which constitute social existence, satisfying Type A requirements is immaterial. An agent who cannot engage in meaningful activities with other agents is hardly human at all.
Given the current endeavor, Type B requirements as social conditions for agency prove most intriguing: to engage in any practice requires that one participates in some (perhaps minimal) shared understanding of the practice in question. This understanding is present in a wide range of mundane and unique cases: in buying a gallon of milk, doing complex mathematics, composing bald assertions, driving on a crowded freeway, and so on. Human behavior can be articulated in theoretical form precisely because it admits of regularities. These regularities are, in most cases, explainable in terms of an understanding the performing agents have of the activities in which they are involved. One knows how to continue counting after 1000 because one has an understanding of what will count as counting.

The second hypothesis mentioned above might be put as follows: some theories exert a causal force in the way human agents view the significance of their actions. Ian Hacking has called this a “looping effect.” This is, undoubtedly, a rather festive occasion given the number of theories circulating in intellectual space, but it is real enough nonetheless. The emergence of clinical psychology provides easy examples aplenty: the category of “obsessive-compulsive,” for example, has led people aware of this category to understand the significance of their actions in new ways. Likewise, the concept of child abuse has led children (and often, thankfully) to understand the significance of events in family life as explainable in terms of a specific wrong-doing on the part of that child’s parents.21

One should not make light of these cases: they are as real as anything else, even if a bit ontologically peculiar. The aim of these remarks is, far rather, to show how certain theories fulfill the Type B requirements stipulated above: an understanding of the significance of one’s actions, or those practices in which one engages, can be altered by theories which call into question that understanding as it stands. The aim of this paper is not to problematize those theories which are marked by a “looping effect.” Rather, it is to suggest one way of approaching theories of human agency in light of this phenomenon. A quick recap is here in order.

A theory of human agency is a set of propositions designed to explain a range of human action. These theories can explain human actions in novel ways, and this explanation can exhibit a looping effect on the agents privy to this novel explanation: it can change the significance of the activities in which agents engage. Because agency has phenomenological social requirements, human agents cannot be agents without understanding their activities in certain ways. This means that a theory about the significance of x, widely disseminated enough, can actually come to constitute the meaning of the practice or activity it aims to explain.

And here is the rub: human practices are articulable in terms of constitutive rules. Human agents understand what will count as engaging in the practice, or activity, in question. A theory attempts to categorize some range of behavior in a specific way by pointing out unifying features of diverse action situations. These theories can thus alter those constitutive rules which would adequately articulate the understanding of the agents involved. It is for this reason, one can argue, that theories of human agency can be viewed as recommendations: they recommend a way of viewing the significance of human actions which can come to constitute that significance, and which is subject to revision.
It is important to limit the scope of the preliminary argument offered above: theories of human agency (often) attempt to paraphrase the ways human beings act: as voters, decision-makers, monogamous partners, and so on, in general terms—they attempt to spell out what is involved in an (appropriate) understanding of human agency; to make articulate the (largely) inarticulate understanding of what counts as human agency. In this sense, it is easy enough to construe them as clarifications of this understanding. But a human agent’s understanding of her own meaningful activity is a multifarious, slippery thing. It is inexhaustibly articulable, if only because the theory in question, via the looping effect, can shape the form of the meaningful activity it aims to characterize. This means that any theory of agency will be incomplete: details will remain on the sidelines, relevant hypotheses in the dustbin.

In another sense, a theory of agency hopes to do too much. Practices manifest a range of behaviors: buying parking passes, doing puzzles, writing poetry, constructing sentences with gerunds, grading papers, planting gardens, rebuilding engines—the range of possible behavior seems, in one sense, endless. A theory of agency attempts to specify this range of behavior by providing a rule (or set of rules) which will determine what counts as human agency. But because the range of behavior is so great, and the agent’s understanding of it infinitely articulable, the theory in question cannot simply clarify—it must also privilege. A theory of agency thus has the following form: A specific range of behavior (x) should count as human agency (y) in a given context (c). The range of behavior in question (the x term) is supposed to be captured by a unifying principle (Kant’s word for “rule”) of some sort: such and such actions manifest rationality, or creative capacities, or a drive for power, or unseen pathologies, or personality traits, or social forces at work on us. Marx characterizes a wide range of practices as demonstrating the creative capacities of human beings, as Nietzsche does with the elusive notion of a “will to power.” These are not to be regarded as empirical theses about human beings. One way to see why this is the case is to notice that they cannot be falsified: no empirical data will show, pace Marxist philosophical anthropology, that human beings are not endowed with a creative capacity which is constitutive of their nature. Likewise, no empirical data can be adduced in support of the claim that there is no such thing as a “will to power.” The notion of falsifiability is crucial to the realm of assertion, and irrelevant to the realm of rules.

But none of this is meant to say that what humans are is subject to the looping effect. An understanding of the biology beneath action—of the skull beneath the skin—will not change that biology, will not alter that skull. The claim endorsed in this paper is a phenomenological one: who human agents are—the way they perceive themselves in phenomenological space— is the product of seeing themselves as beings capable of performing certain sorts of actions and living certain sorts of lives. A theory of human agency (and not, say, a theory of humans qua biological beings) will attempt to make sense of this existence in terms of law-like regularities: human agents have a capacity for autonomy, are rational/speech animals, are beings with a creative capacity from which they are alienated in a capitalist society, and so on. And this activity can come to constitute the significance of some range of behavior by altering what is seen as a truly worthwhile life. It is this fact, to state the conclusion of the argument given above bluntly, which justifies viewing theories of agency as recommendations.
V. Conclusion

The argument of the preceding section is baptized "preliminary" for good reason. The aim of this article has been to offer some reasons for thinking that one should approach theories of agency in a more "therapeutic" fashion (though there is certainly danger in this word). If it is true that theories of this sort, when widely disseminated enough, can constitute the way situated human agents regard the significance of their actions, then theorizing about agency has an ethical dimension. The argument of the preceding pages attempted to highlight this dimension by distinguishing three types of statements, and attempting to show that a theory of agency falls into the third of these (albeit in what is perhaps an inadequate way). But much more remains to be said. The aim of this preliminary sketch was simply to provide one way of talking about such matters—one that no doubt might best be improved by the insights and criticism of others.

University of California, Riverside

NOTES


2. This is not a defense of the sense/reference distinction (though, admittedly, it seems wholly unproblematic). This distinction is meant to outline the features of an assertion. The same point can be made (and, indeed, has been made) by other means.

3. This is not to say that a statement which postulates some entity, say, without empirical evidence, fails to be assertoric. The point is rather that the statement must be in principle verifiable or falsifiable. This is a much weaker claim than saying that an assertoric statement must already be verified or falsified.

4. This is precisely what Wittgenstein called "nonsense."


6. Searle insists that he is not engaged in phenomenology, but in logical reconstruction (see his response to Dreyfus on this point). If one construes phenomenology as a subject-centered approach to philosophical questions, this is undoubtedly true. The significance of the y term has nothing to do with the way an individual subject understands it. Rather, a constitutive rule counts as a rule because all understand it in (roughly) the same way.

7. The statements here under consideration are those where the articulation of a constitutive rule is an accurate one. Because it is possible to articulate these rules in myriad ways, there will obviously be many cases in which one will want to problematize someone's articulation. The point here is that the rule cannot be falsified empirically, though it can be articulated inadequately.

8. This is not to say that empirical considerations never mattered. Grammar, in its infancy, had to consult actual language use in order to articulate its constitutive elements (much as one would consult the actual proceedings of some practices in order to be able to articulate constitutive rules regarding it—though one would not always need to, as in the case of, say, legal practices). Nevertheless, (b) is not empirical—even if at one point it was. Empirical claims can become constitutive rules: (b) clarifies what it means for something to be a sentence. It defines a domain by providing a rule. It is thus, in one sense at least, analytic.
9. This will receive further treatment in the following section.

10. More explicitly: One cannot say "Unless one is driving on the right side of the road, one is not driving at all."

11. This will be more fully explored below.

12. This, of course, is the thesis of Milgram's influential *Obedience to Authority*.

13. The "more or less" is crucial here. No personality theorist would deny that traits can change during the course of one's life—but it is the regularity of such traits, in general, which marks them as explanatory. If one were to manifest different traits every other second, these traits would hardly be a useful means for predicting and describing one's actions. It has also been suggested in a number of places that the stability of traits increases with age. See, e.g., R. McCrae and P. Costa's *Personality in Adulthood* (New York: The Guilford Press, 1990).

14. The Principle of Construal is found in a host of places in psychology. See Piaget's "schema," for instance, as well as recent uses of this in personality theory (e.g., Harter, 1999).

15. Questions about legal precedent fall nicely into this category: a judge deals explicitly with recommended rules. Should this case count as subsumable under these laws? His answer makes a recommendation (in many cases, that of either the defense attorney or the prosecutor) into a constitutive rule (another precedent).

16. This is precisely the value of much of the (good) social constructionist literature: it allows one to alter one's perception of some phenomenon from "x counts as y" to a statement of the form "x should count as y," where this statement is seen as the product of a specific history without necessity. When the "counting as" relation has been shown to be (politically) problematic, one can deny that the "should" has any force.

17. Notice that the words "theory" and "theater" have the same etymological root: the Greek "thea," from which both of these terms probably derive, means "viewing."

18. It is here assumed that there is no such thing as *unsituated* human agency: acting as agents always occurs in some environment.

19. This term is Ian Hacking's. Explication is forthcoming.

20. Explanations of neuroticism often indulge the view that requirements of Type A have not been met. In more extreme cases, development outside of a social setting atrophies completely. This is clear from the heinous cases the social sciences have brought to attention: a child locked in a closet for the first twelve years of her life is not recognizably human; a child spending the first formative years of his life in the wilderness cannot acquire linguistic competence, and has trouble conforming to the mores of a social world.

21. The scope of this agreement will vary, of course, from case to case.

22. See *The Social Construction of What?*

23. On the other hand: many are the distraught parents who have heard a spoiled suburbanite proclaim, after a minor belting, that he would "turn them in for child abuse."

24. Thus, if one were so inclined, one could articulate theories of agency as embodying two constitutive claims: some range of behavior (x) counts as displaying some principle (y), and this principle (x) should count as human agency (y).

25. The above examples might as easily be called "theories of human nature." The term "human agency" has nonetheless been retained, despite the connotations, because it emphasizes a correlation between what humans are and what they do.
LIST OF REFERENCES


