Brandom's Theory of the Institution of Norms

Leonardo Marchettoni

Abstract: In this paper I raise a number of issues concerning Brandom's pragmatist explanation of norms. I will argue that Brandom's attempt to explain normative statuses through recourse to normative attitudes does not succeed in distinguishing norms from regularities of behaviour. I suggest that talk about normative attitudes is translatable into dispositionalist terms, within a language devoid of normative notions, and that the thesis of the institution of norms by the practical attitudes of the members of the community fails to make sense of the idea of objective normative statuses existing above what single practitioners hold as correct according to their understanding of norms. In the first section I will consider Brandom's discussion of the rule-following problem presented in the first chapter of *MIE*; in the second section I will analyse Brandom's arguments against the accountability of attitudes in a non-normative language, and put forward a dispositionalist reading of Brandomian semantics; I will then try to consider whether the dispositionalist reading of normative attitudes entails a corresponding naturalisation of Brandom's semantic project, and examine Brandom's critique of the tenets of AI functionalism in *Between Saying and Doing.*

Keywords: Robert Brandom; inferentialism; rule-following; norms; dispositions; pragmatism.

My aim in this paper is to discuss a number of issues concerning the theses of the irreducibility of normative attitudes to non-normative notions and the institution of norms by attitudes, which Brandom defends in *Making It Explicit* (Brandom 1994, henceforth *MIE*). The basic suggestion is that the pragmatist strategy of explaining the normative aspects of intentional phenomena does not succeed in distinguishing itself from a dispositionalist approach: pragmatism about norms explains normativity by recourse to items that are, in the end, indistinguishable from behavioural dispositions, despite Brandom's assertion to the contrary. I suggest that talk about normative attitudes is translatable into dispositionalist terms, within a language devoid of normative notions, and that the thesis of the institution of norms by the practical attitudes of the mem-

bers of the community fails to make sense of the idea of objective normative statuses existing above what single practitioners hold as correct according to their understanding of norms.

The plan of the essay is as follows: in the first section I will consider Brandom's discussion of the rule-following problem presented in the first chapter of *MIE*; in the second, I will analyse Brandom's arguments against the accountability of attitudes in a non-normative language, and put forward a dispositionalist reading of Brandom's semantics; I will then try to consider whether the dispositionalist reading of normative attitudes entails a corresponding naturalisation of Brandom's semantic project, and examine his critique of the tenets of AI functionalism in *Between Saying and Doing* (Brandom 2008, henceforth *BSD*).

1. *Brandom on Rule-Following*

In the first three chapters of *MIE*, Brandom introduces the basic elements of the overall approach pursued throughout the book. These elements are: *the normative nature of intentional phenomena* – that is, the idea that one of our principal features as rational beings is the ability to use concepts in order to represent things in the world, and that the items through which we try to achieve this result can be classified as *correct* or *incorrect* representations depending on whether or not a concept is *properly* employed in them; secondly, *the inferential structure of conceptual content* – the suggestion that the activity of representing things and states of affairs in the world is connected with an ability to pass from one representation to a number of others through *inferential links*, which articulate the concepts employed. The final element is *the pragmatic conception of semantics* – the idea that grasping the conceptual content of our assertions requires knowing what role each assertion plays in the context of the global dynamics of our linguistic practice.

Put together, these three elements form a comprehensive image of what Brandom calls *sapience*.

Sapience of the sort distinctive of us is a status achieved within a structure of mutual recognition […] The specifically *discursive* character of that normative social structure […] consists in the *inferential* articulation of those recognitive practices (*MIE*: 275, italics in original).

Sapience is what distinguishes human beings from merely sentient beings. It can be expressed as a kind of subjection not only to natural laws but also to inferentially articulated norms. But this does not mean that sapience is a kind of abstract essence that only agents of a given kind possess. On the contrary, Brandom

stresses that the possession of sapience is rather a matter of exhibiting certain practices that, in principle, can be shared by any agent (see also Brandom 2010).

The assumption of this normative starting point calls for an extensive discussion of the nature of norms and of the traditional problem of rule-following. This task is accomplished in the first chapter of *Making It Explicit*. After a general introduction to (the theme of) the normative significance of intentional states, in which Brandom traces his explanatory strategy back to Kant, Frege and Wittgenstein, the core of the chapter is devoted to outlining a theory of norms capable of addressing the rule-following paradoxes. The difficulty here is clearly in explicating how the act of behaving in accordance with norms can be accounted for in a way that preserves the existence of facts regarding what counts as the correct application of a rule.

The plan of the argument follows a well-travelled route. First, the "regulist" (Platonist) view, according to which the normativity of intentionality presupposes the existence of explicit rules, is criticised. The reference here is obviously to the passages in the *Philosophical Investigations* in which Wittgenstein shows how the Platonist conception entails the unfortunate consequence of an infinite regress of interpretations (*MIE*: 18-26).[1] Brandom then considers the "regularist" theory, that is, the view that explains the existence of norms through regularities of behaviour or dispositions towards action. The problem with this proposal is that any finite sequence of behaviour is consistent with an infinite number of possible continuations, and thus that behaviour might accord with an infinite number of rules; this makes it impossible to distinguish between correct and incorrect performances (*MIE*: 26-29).

The upshot of these first two steps suggests that, in order to make sense of the idea of the normative character of intentionality, we need an account of norms that steers a middle course between the Platonist and the regularist views, that is, an account that replaces explicit norms with norms *implicit* in practice. However, while Kripke and others (Kripke 1982; Wright 1980) assume that the failure of both Platonist and regularist accounts brings us to the conclusion that the normative character of rule-following can be explicated only in the context of communal practice, Brandom firmly rejects this inference. In Brandom's view, the thought that the performances a member of the community produces can be properly assessed as appropriate or inappropriate only by the community to which the individual belongs lays itself open to two orders of objection.

The first criticism concerns the fact that "communal assessment theorists have a tendency to personify the community" (*MIE*: 38). The problem is that it

---

[1]   Another important source of inspiration is Sellars 1954.

is not clear how a community as a whole can assess or endorse the practices of its members. If "universal agreement is too much to ask", how can we decide what kind of response counts as an endorsement and which counts as a critical appraisal? The verdict can, no doubt, be deferred to the experts in the field, but, as Brandom recognises, this division of labour involves a normative notion: someone is an expert because she has a certain authority. So it seems – and this is the second line of criticism – that the communal assessment theorist, in so far as she avails herself of notions like "expert", "authority", etc., turns out to be guilty of "smuggl[ing] normative notions illicitly into what purports to be a reductive, non-normative regularity theory" (*MIE*: 38).

The approach followed by the communal assessment theorists, as they were known, proved as flawed as regularity theories. In both cases, there is a problem concerning the possibility of distinguishing between correct and incorrect performances without appealing to normative notions. These failures suggest that there is no way to express the normativity of intentionality without referring to norms (*MIE*: 45-46).[2] We must therefore accept that normative vocabulary is non-reducible and non-replaceable. It is certainly important, however, to understand exactly what consequences the idea of the non-reducibility of normative notions entails. At this point Brandom owes us a detailed account of the idea of non-reducibility of normative notions, explaining how rules are related to regularities of behaviour, as described in naturalistic terms.

Brandom elaborates his alternative by distinguishing two claims about the emergence of norms from non-normative facts. The first claim asserts that "unlike natural properties, normative proprieties are in the eye of the human beholder" (*MIE*: 47). This insight – that values are produced by us – runs counter to a stronger claim. This is the physicalist

conviction that the specification of the values of an appropriate range of dynamic variables for all the fundamental particles provides a *complete* description of everything that deserves to be called *real* (*MIE*, *ibidem*, italics in original).

Brandom's crucial move consists in maintaining that, if we accept the first claim and reject the second one, we can attain an explanation of norms that clarifies our status as rule-makers without assuming a full-blooded reductionist stance. Indeed, taking it for granted that the institution of norms is a consequence of our judgments and assessments of value, and hence cannot be accounted for in a language devoid of normative terms, does not entail these judgments and assessments being themselves describable in "purely physical terms".

Following this path, Brandom is led to an explanation of normative facts

---

[2]   Brandom quotes extensively from McDowell 1984 in support of this claim.

that takes normative statuses of performances – i.e. statuses of performances in relation to their correctness or incorrectness – as supervenient on practical attitudes. Normative statuses are taken to be instituted by our practical attitudes, that is, by our treating certain performances as correct or incorrect. This dependence assures us that, where we encounter the same attitudes, we will also find the same statuses. At the same time however, since attitudes are not describable in naturalistic terms, the stronger claim that normative statuses supervene on merely natural facts does not follow.

At this point we have gathered all the elements needed to evaluate Brandom's solution to the problem of rule-following. The basic idea is that, if we take normative statuses of performances to be supervenient on normative attitudes, explaining the former with reference to what we treat as correct or incorrect without endorsing any kind of reductionism of the latter, we can steer a middle course between regulist and regularist temptations. This approach, while at first sight promising, is ultimately unconvincing, as I will now show.

2. *Attitudes and Dispositions*

Brandom's solution to the problem of rule-following centres on the idea that we can explain the existence of rules if we focus on our activity of treating performances as correct or incorrect. In this view, normative statuses supervene on normative attitudes, which in turn are deemed to be non-describable in purely naturalistic terms. Is this assumption reasonable? To answer this question, we need to examine his account of normative attitudes.

Normative attitudes are assessments, "assignments to performances of normative significance or status, as correct or incorrect according to some norm" (*MIE*: 35). But assessments can be understood as dispositions to sanction, that is, to reward appropriate and punish inappropriate performances, as Brandom recognizes (*MIE*, pp. 34ff).[3] It seems, therefore, that normative attitudes are completely explicable in naturalistic terms. They seem reducible to clusters of behavioural dispositions. When we talk about someone's attitudes, we are actually talking about their disposition to react to certain performances, sanctioning them, whether positively or negatively. But if this hypothesis were correct, we could conclude that there is a naturalistic description of norms that is couched entirely in non-normative terms.

The issue of the reducibility of attitudes to naturalistic reports via behavioural disposition met with a strange fate indeed. Some commentators rashly

---

[3]   Brandom is probably reminiscent of Sellars 1974, who spoke of "constraints" and "incentives".

reject the thesis that normative attitudes are reducible (Wanderer 2008: 14-17; Peregrin 2012: 88-90) whilst others endorse it without discussion. One such is Anandi Hattiangadi, who writes:

It is unclear how Brandom's view differs from a straightforwardly naturalistic one […]. So put it appears as though Brandom is offering a dispositionalist account of the determination of correctness – since the starting point includes nothing more than behavioural dispositions. Moreover, […] nothing is added that would distinguish the account from dispositionalism (Hattiangadi 2003: 424-425).[4]

But in *MIE*, we can find a two-stage argument against the reduction[5] of attitudes to non-normatively specifiable dispositions, so it seems that, if we want to settle this issue, we should start discussing these passages. At the outset, however, it is fair to say that the attitude Brandom displays towards dispositional jargon has changed since the publication of *MIE.* In fact, while he criticized dispositionalism in his *magnum opus* by way of countering naturalistic reductionism, he seems to adopt a more liberal stance in his later works, sometimes employing dispositions as an alternative way of accounting for normative notions.[6]

The first part of the argument is based on the observation that the normative character of the meta-language in which norm-instituting social practices are specified is irreducible to naturalistic accounts:

it is important to realize that it is one thing to understand practical assessment as sanctioning, and quite another to understand sanctioning in non-normative terms such as reinforcement. … Defining normative attitudes in terms of dispositions to apply sanctions does not by itself reduce the normative to the non-normative – it just trades off one sort of norm for another. At the most basic level, to reward someone is to offer some good … and to punish them is conversely to inflict something bad. Benefit and harm, desirable and undesirable, are concepts that also have normative senses. Indeed, these senses would seem to be primary, so that some sort of reductive hypothesis would be needed to naturalize them (*MIE*: 42).

According to Brandom, the reduction of attitudes to dispositions to apply sanctions cannot be a proper reduction of normative notions to non-normative ones, because the description of attitudes as dispositions to sanction is not entirely couched in naturalistic vocabulary. In fact, the "sanction" concept is a normative notion, something that refers back to a normative theory of what

---

[4]   See also: Grönert 2005: 163-164. For a critique of Hattiangadi, see Peregrin 2012: 88-90.

[5]   Reduction is to be intended as the activity of eliminating, "in favour of non-normative or naturalistic vocabulary, the normative vocabulary employed in specifying the practices that are the use of language" (*MIE*: xiii).

[6]   See, in particular, Brandom 2011; but see also the recourse to dispositionalist jargon in *BSD*.

benefit and harm consist of. Therefore, in order to naturalise it, one would need a reductive hypothesis making it possible to translate good and evil into naturalistic idiom.

However, I will contend that, even if Brandom is correct in maintaining that the meta-language in which we describe a sequence of behaviour as an instance of reward or punishment – that is, conduct that is performed with the specific intention of sanctioning, positively or negatively, in accordance with a given set of norms – must contain normative notions, the proper individuation of the relevant behavioural responses does not require recourse to normative notions. Indeed, since individuals who share the same conception of benefit and harm may disagree in their attitudes towards a given performance and, conversely, subjects having different conceptions of good and evil may express the same attitudes, the choice of the sanctions is not relevant to defining the status of a performance as correct or incorrect.[7] Moreover if this were not precluded by the variability of sanctions, we could express the behavioural outputs in which normative attitudes result, as dispositions, in purely non-normative terms. For example: we can imagine a context in which people are disposed to adopt a limited number of behavioural responses – think of a stream of cars in which drivers express their disapproval of certain conduct on the part of other drivers by sounding their horns. In such a case, a dispositionalist account of normative attitudes entirely couched in non-normative terms would be available. Such an account may take the form of a collection of sets, containing, for each driver, a non-normative specification of the conduct in response to which that particular driver is disposed to sound his or her horn. For example: Driver j sounds her horn when another driver attempts to overtake her, etc. The use of normative concepts like benefit and harm adds further determinations, introducing new classifications into the "basic" naturalistic description. However, the specific behavioural responses that actualise the activity of sanctioning can be individuated in non-normative terms. For this reason, normative notions are not eliminable from the meta-linguistic accounts, not because they make an implicit reference to the normative conceptions of the evaluator, but because when describing a course of action as an instance of sanctioning, one is using normative vocabulary.

---

[7]   It is useful to compare this case with that of communal assessment theories about rule-following. Brandom maintains that communal assessment theorists make illicit reference to a normative concept when they define correctness in terms of another normative concept, the concept of "expert" (*MIE*: 39). This objection seems reasonable, because referring to the normative notion of "expert" is necessary in order to determine the extension of the set of correct performances, since this set is by hypothesis composed of just those performances that are so evaluated by experts. So in this case – but not in the case of the reduction of attitudes to dispositions to sanction – reference to a normative concept is *necessary* for the purposes of determining the content of attitudes towards performances.

This circumstance reinforces the suspicion that Brandom's idea of accounting for the social institution of norms in a normative language that reinterprets a sequence of moves describable in non-normative terms should address the existence of a gap between the level of norms and that of naturalistic behaviour. After all, Brandom owes us an account of how to move on from the naturalistic level to the level of norms: How is one to decide which behaviour counts as an appropriate response to a given norm? The non-normative description does not include the conceptual resources needed to solve such a problem, whereas, any account given in normative terms would prove irredeemably circular (Rosen 1997: 167).

In order to avert qualms of this kind, Brandom sets out to show, in the second part of the argument, that when the social construction of norms becomes increasingly complex, it is not even possible to offer a non-normative individuation of the behaviour through which the activity of sanctioning is performed. He starts by noting that positive and negative sanctions need not consist of rewards and punishments. Indeed, they "may consist of acclaim and censure that itself has only a normative significance" (*MIE*: 43). A correct performance can be rewarded by release from an obligation; in the same way, license might be withheld to punish an incorrect execution. In such cases, there is no direct shift from normative evaluation to bestowal of benefits or imposition of damage; we simply face "a change in normative status rather than natural state" (*MIE*: *ibidem*). This means that we may distinguish between *external* and *internal* sanctions. External sanctions are those that are expressible in non-normative terms, like offering food or beating with sticks, whereas internal sanctions are those that involve only a change of normative status. Now, for Brandom,

[i]t is possible to interpret a community as instituting normative statuses by their attitudes of assessment, even though each such status that is discerned is responded to by sanctions that involve only other normative statuses. … Such an interpretation would not support any reduction of normative status to non-normatively specifiable dispositions, whether to perform or to assess, whether individual or communal (*MIE*: *ibidem*).

I find this part of Brandom's argument unconvincing, because a dispositionalist account of sanctioning, even internal sanctioning, is indeed available, as I will try to show. First, however, since talk of dispositions almost immediately conjures up scenarios in which the activation of a given disposition is prevented by some countermeasure, it will be helpful to say something in advance about the ideas talk of dispositions should encapsulate. I will take dispositions to mean intrinsic resources that make a given agent capable, in a given situation, of performing a given task in a given way. In other words, I am assuming that the dispositions we are considering – dispositions correlated to normative

attitudes – are intrinsic and canonical. They are canonical because normative attitudes are assessments of performances, and intrinsic because what counts as a performance, and the ways in which it will be assessed, depend on subjective standards that may themselves be specified by a further disposition.[8]

To overcome the problem bound up with the distinction between external and internal sanctions, we might note that once an internal sanction is applied, it is reasonable to expect that the community members will begin to adjust their behaviour in accordance with the change of normative status that follows the sanction: for example, by declaring what additional duties the transgressor must now fulfil. Attributing a behavioural disposition is basically a way of associating a certain behavioural output with a given circumstance without appealing to inner states, therefore it is still possible to conceive of normative attitudes pertaining to internal sanctions in dispositional terms, not as dispositions to sanction, but as sets of dispositions to calibrate one's behaviour as a consequence of changes of normative status. And, for what I said above, it is possible to individuate the "proper" sets of dispositions without making use of normative notions. Therefore, the possibility of internal sanctions does not jeopardise the availability of a non-normative description of the empirical consequences of violation of and compliance with norms.

Let us consider another example. Instead of a context in which all individuals are disposed to sanction in the same way, we may imagine a community in which every member must perform some task from a given set *A*. When a practitioner fails to carry out her duty, her normative status changes, and she is assigned to performing a task from another set, say *B*. If she fails again, she is assigned a task from a third set, *C*. Eventually, if she persists in her negligence, the community member is subjected to an external sanction *S*. Were we to describe the normative attitudes of these folks in evaluating other community members' performances, we could resort to dispositionalist jargon as follows: everyone has a set of *first-order* and *second-order* dispositions. First-order dispositions are instructions that specify how to accomplish certain tasks and how to behave towards other community members, and second-order dispositions are instructions that specify how to react to other practitioners' performances. If we assume that each individual has some default dispositions and define a second-order disposition as a disposition that alters one's own pre-existing dispositions, we can interpret normative attitudes as second-order dispositions, i.e., dispositions to keep track of others' performances and modify one's own default dispositions in accordance with new evidence concerning how a certain performance has been carried out.

---

[8] See Choi, Fara 2012 for an explanation of the relevant terminology.

In the end, the predicament is not so different from that of the first example: normative notions perform an essential function in unifying a given set of conduct, by linking it to a system of norms. This does not mean that a non-normative listing of those instances of conduct is not available.

The original behaviouristic motivation behind the dispositionalist interpretation of attitudes can thus be preserved without reintroducing normative notions. Since there is no causal role a suitable choice of dispositions cannot capture, we may expect to translate Brandom's account of normative attitudes into dispositional terms. For these reasons, I think that a dispositionalist, non-normative reformulation of normative attitudes is available. Indeed, once a convenient set of first-order dispositions and rules that explain in which way an agent alters her first-order dispositions as a consequence of acquiring new information has been defined, it becomes possible to display a purely dispositional account of the activity of assessing performances in which changes of normative status are expressed in terms of second-order dispositions. In this account, the difference between internal and external sanctions does not matter, because it corresponds to a difference in the sets of second-order dispositions.

It is noteworthy that Brandom himself adopts a similar approach in several places in order to explain the normative role of recognitional abilities. For example, think of the story of the "queen's shilling" in the third chapter of *MIE*. Brandom writes:

The significance of taking the queen's shilling lies in its being an undertaking of a commitment on the part of the recipient, altering the attributions of commitment by those who appreciate the significance of the performance. It entitles other authorities – those who according to the antecedent score already had undertaken various commitments or duties and entitlements or sorts of authority, those who therefore play a certain role or hold a certain office in the system of practices in question – to punish the performer in particular ways under particular circumstances. The normative significances of performances and the deontic states of performers are instituted by the practice that consists in keeping score by adopting attitudes of attributing and acknowledging them (*MIE*: 166).

It becomes evident from the description how a change in normative status may involve modifications expressible in terms of dispositions to behavior. And in a later essay he writes:

My taking *your* K-response to have been authorized by a K-desire that serves as a standard for the success of your K-taking, and taking that K-response to have been *correct* or *successful* by that standard, is my acknowledging the *authority* of your K-taking, in the practical sense of being disposed *myself* to take *as* a K the thing *you* took to be a K. Taking it that the kind of fruit you ate really was food, in that it satisfied *your* hunger,

is being disposed to eat that kind of fruit myself when and if *I* am hungry, i.e. have a *desire* of the same kind. This is a second-order disposition, involving a change in my first-order dispositions (Brandom 2011: 41-42, italics in text).

Second-order dispositions explain the normative role of relations of authority between recognizers, in virtue of which each agent acknowledges another agent's desire and activity as a standard against which to evaluate the correctness of a given performance. Moreover, according to Brandom, second-order dispositions, as changes of first-order dispositions, may account for the transitive character of 'robust recognition', that is, recognition in which the recognizer acknowledges the authority of other agents' recognitions over her own.[9]

It is obvious that the foregoing sketch cannot be taken to provide a complete account of normative attitudes in dispositionalist terms,[10] but the purpose of the preceding discussion is not to offer a new example of reduction of normative attitudes. The main problem I would like to highlight is that Brandom's treatment of the hypothesis of reducing normative attitudes to dispositions is inadequate. As I have explained, Brandom's criticism takes into account only a very rough reductive hypothesis: the hypothesis of interpreting normative attitudes as dispositions to sanction. But the example of first-order and second-order dispositions shows that there may be dispositional approaches that cannot be dismissed so quickly. Failing a general argument against the explanation of attitudes in terms of dispositions, and given that the indispensability of normative notions has not been proven, Brandom's solution to the problem of rule-following does not succeed in drawing a via media between normative and non-normative accounts.

At this point, however, one might wonder why Brandom has not explicitly criticised the behaviouristic suggestion that underlies the approach to normative attitudes in terms of dispositions. After all, Brandom could have directly attacked the idea according to which attitudes are analysable in terms of dispositions, instead of raising doubts about the accountability of sanctions in non-normative terms, following, for example, Geach's lead (1957). We have seen how Brandom tries to undermine the non-normative reduction of normative attitudes by noting that, at least in certain cases, normative attitudes are not analysable in terms of disposition to punish or reward. He does not, however, attempt to discredit the more general thesis that attitudes are analysable in terms of behaviour. I think that this apparent oversight is connected with the comprehensive design of his pragmatist strategy. Indeed, Brandom cannot afford to

---

[9]   Brandom 2011: 44. See also Testa 2011: 300-302 on these topics.
[10]   Developing a better model may require introducing *higher-order* dispositions. But this is a problem that does not affect my argument.

reject the behaviouristic principle whereby attitudes are analysable in terms of behaviour, because it is on this very principle that the fundamental aim of *MIE* – to offer an account of intentional phenomena as phenomena ruled by norms and of norms as a product of both our *activity* and the social interplay between members of a community – is based. Brandom wishes to combine his pragmatist inclination with an anti-reductionist stance that preserves the autonomy of normative vocabulary, thus constraining the behaviouristic principles that lie beneath his pragmatist inspiration. If my criticism is correct however, he fails to construct a sensible, non-reductionist brand of behaviourism.

Moreover, it is not at all difficult to adapt the dispositional model to the case of linguistic practice. Such an endeavour becomes easier because, as a recent commentator writes, Brandom "accepts … that practical attitudes can be highly complex strategies of interaction with the world" (Kiesselbach 2012: 105, No. 6). This assumption is visible in the way Brandom himself builds up his scorekeeping semantics, in which the recourse to normative notions is mixed with an ingenious model in which inferential relations are mimicked by practical attitudes. In order to rephrase semantic items in dispositional terms, inferential as well as representational and practical dimensions of linguistic activity must be taken into account. Bearing it in mind that practical attitudes can be expressed in non-normative terms, we may do so as follows:

For a given sentence φ, the meaning of φ, M^φ, consists of an ordered quintuple, M^φ: $<C^\varphi, E^\varphi, I^\varphi, A^\varphi, P^\varphi>$, where C^φ is the set of commitment-preserving inferences in which φ (along with other auxiliary hypotheses) plays an essential role as premise, E^φ is the set of entitlement-preserving inferences in which φ (along with other auxiliary hypotheses) plays an essential role, I^φ is the set of the sentences that are incompatible with φ, A^φ is the set of the states of affairs of which φ constitutes an appropriate observation report and P^φ is the set of actions to which the assertion of φ commits the speaker.[11]

Each speaker in her practice of language follows what she believes to be the conditions of correct usage. So she must have some idea of what further claims she is committed to as a consequence of her assertion of φ, what assertions she is entitled to make as a consequence of the entitlement to φ, what sentences are incompatible with φ, what states of affairs can be described by means of φ and

---

[11]   See *MIE*: 188-198. Some of these sets may be infinite, for example: the set of states of affairs of which the sentence "the table is red" constitutes an appropriate observation report. Such a consequence depends, however, on the decision to ground normative attitudes on reliable differential responsive dispositions that can discriminate between an infinite number of cases. So, if there is a problem with infinite sets, it is a problem that does not depend on the specific features of my formalization.

to what actions she is committed as a consequence of her assertion of φ.[12] In this formalization, each reference to normative notions is replaced by enlisting inferences, sentences, states of affairs, actions, and so on.

Two or more speakers may, however, differ in their judgments about what inferences a given utterance licenses or about what states of affairs a given sentence can correctly describe or, indeed, to what actions their assertions commit them. This complication is acknowledged by Brandom himself when he recognises the perspectival character of conceptual content, i.e., the fact that what constitutes the content of a sentence varies from speaker to speaker in accordance with the linguistic customs each speaker has acquired *(MIE:* 185). The awareness that content is perspectival calls for a slight reformulation of our previous definition of meaning in terms of sets of sentences and states of affairs. Since each speaker may attribute a different meaning to a given assertion, it appears reasonable to index meaning in relation to speakers. The result we obtain is that the meaning of a given sentence φ is expressible as an indexed family of sets, $M_i^\varphi{:}{<}C_i^\varphi, E_i^\varphi, I_i^\varphi, A_i^\varphi, P_i^\varphi{>}$, in which each M_j^φ represents the meaning of φ from the perspective of a single speaker j.[13]

Once we have defined what performances are relevant for classifying correct or incorrect uses of sentences, it is possible to state the content of an inferentialist approach to semantics more accurately. Normative statuses pertaining to semantic norms are connected with selecting, for a given sentence φ, and for a given speaker j, the inferences that form sets C_j^φ, E_j^φ,[14] the sentences that form set I_j^φ, the states of affairs that are collected in set A_j^φ and the actions that belong to set P_j^φ. The upshot is that we can easily transpose the abstract analysis of attitudes onto semantic norms if we take into account the five components of linguistic meaning, ${<}C_j^\varphi, E_j^\varphi, I_j^\varphi, A_j^\varphi, P_j^\varphi{>}$.

Consider, for example, a speaker j and a set of sentences S. Since content is perspectival, for each sentence $\varphi{\in}S$, there is a quintuple M_j^φ, which describes the conditions of correct usage for that sentence from the speaker's perspective. M_j^φ indicates which inferences follow from φ and which states of affairs can be reliably reported by means of it from j's perspective, and so this ex-

---

[12]   Note that the set of commitment-preserving inferences in which φ serves as the conclusion can be recovered from $I^{\neg\varphi}$, since $\neg\varphi$ is incompatible with ψ iff $\neg(\neg\varphi\wedge\psi)$ and hence $\psi\rightarrow\varphi$. See Brandom 2009: 46. See also *BSD:* Ch. 5. In order to extend this model to subsentential expressions, substitutional and anaphoric commitments would have to be taken into account too. This is beyond the scope of this essay.

[13]   The perspective of a single practitioner must not be confused with a context of use. Semantic content varies from a context to another according to the collateral commitments available as auxiliary premises to practitioners. See *MIE:* 139.

[14]   One should note at this point that the sets of auxiliary premises associated with C_i^φ and E_i^φ, too, may vary according to the speaker. This fact is decisive for the adjustment of usage to contexts. However, I must ignore this complication.

plains how j is prepared to use φ in her linguistic exchanges with other speakers. We can now call $M_j^S = \{M_j^\varphi \mid \varphi \in S\}$ the set consisting of all the quintuples M_j^φ, for all the sentences included in S. Now we can extend this rough model in a simple way. We can imagine that, when entering into contact with another speaker k, j is likely to behave linguistically according to the conditions stated by M_j^S and will also evaluate k's utterances along the same lines. This last condition implies that j will keep track of k's entitlements and commitments as well as her own, and will form corresponding deontic attitudes that specify which further performances k is committed or entitled to, thus regulating further linguistic behaviour. Moreover, it requires that j will note the divergences between k's further utterances and her expectations and revise her future expectations accordingly.[15] If there is any reason to recognise some special linguistic authority in k, j will modify her attitudes to linguistic behaviour and evaluation. Recalling the previous discussion about the dispositionalist interpretation of attitudes, we may attempt to formulate this point in terms of second-order dispositions. If we interpret the set M_j^S as determining j's default first-order dispositions to linguistic behaviour, we can conceive of deontic attitudes as second-order dispositions that specify how j will keep score of k's behaviour. This may be by addition of, for example, the further sentences to which k is committed or entitled as a consequence of his assertions, deletion of the sentences that are incompatible with them, and possibly, indeed, revision of her previous dispositions in accordance with new evidence concerning how a certain linguistic performance has been carried out.[16]

3. *Dispositions, Naturalism, and AI Functionalism*

So far, I have argued that the arguments Brandom sets out in *MIE* against the reduction of norms to non-normative vocabulary are not completely convincing and that, starting out from the inferentialist semantics developed in *MIE*, it is possible to sketch the main lines of a dispositionalist approach to linguistic practice. Now it is time to consider more closely what kind of ramifications develop from this reading.

The main problem I would like to consider in this section is whether a

---

[15]   This means that j will keep two separate registers, concerning, respectively, the objective meaning of k's assertions – objective from j's perspective, obviously – and the meaning that k himself attributes to his assertions.

[16]   It is reasonable to imagine that there should be a set of dispositions that specify in which cases the speaker has to recognize in her interlocutor some kind of linguistic authority, but this is a point I will not pursue further. Meaning displays its dynamic character in this possibility of revision.

dispositionalist reading of Brandom's inferentialist semantics entails a naturalisation of his pragmatist strategy. It is obvious that the answer given to this question depends on the theory of dispositions one endorses. If one thinks that dispositional properties can be reduced to categorical properties expressible in naturalistic language, the possibility of defining attitudes in dispositional terms implies the naturalisability of inferentialist semantics.

Things change if we adopt a different account of dispositions. For example, if we adopt Stephen Mumford's functionalist theory of dispositions, the road to naturalism is not so straightforward. According to Mumford, the dichotomy between categorical and dispositional properties does not reflect an ontological dualism. On the contrary, the categorical and the dispositional idiom are two ways of talking "about instantiated properties in the world" (Mumford 2003: 192). The underlying ontological assumption is a kind of "neutral monism", that is, an ontological stance according to which dispositional and categorical tokens can be identical without dispositional vocabulary being reducible to categorical vocabulary and vice versa. In this context, an ascription of a dispositional property is nothing more than the ascription of the ability to mediate causally between stimulus events and manifestation events.

If we follow Mumford's functionalist theory of dispositions, it is clear that the prospects for a naturalisation of attitudes are less promising. If dispositions are not reducible to naturalistic properties, translating Brandom's scorekeeping model into dispositional terms serves no purpose as far as achieving a naturalisation of his semantic project is concerned. The dispositionalist jargon is a kind of empty box that cannot carry a specific ontology inside it. To say that a given subject is furnished with a given set of linguistic dispositions is to say nothing about the way those dispositions are realised. There is, therefore, no direct shift from rephrasing attitudes in dispositional terms to arriving at the conclusion that they are accountable in naturalistic vocabulary. Construing Brandom's position as a kind of naturalism would require us to discuss the effect of reducing behavioural dispositions to naturalistic items. I am not sure such an argument is indeed available; therefore, I will dub Brandom's approach a form of *quasi-naturalism*.

This conclusion has obvious consequences for the rule-following problem Brandom aimed to solve in the first chapter of *MIE.* Brandom's solution centred on the proposal of a middle way between regulist and regularist approaches, capable of preserving normativity without falling prey to infinite regress. Now, reinterpreting the idea of the institution of norms by normative attitudes of the speakers in dispositionalist terms introduces a further move in this dialectical exchange. Indeed, as we have seen in the previous paragraph, conceiving of meaning in dispositionalist terms and, at the same

time, acknowledging the perspectival character of conceptual content induces a reading of deontic attitudes as second-order dispositions. But if we admit the variability among speakers of the first-order dispositions that determine perspectival conceptual contents, there is no reason not to allow for a corresponding variability in the set of second-order dispositions that constitute deontic attitudes. It follows that there is no room for objective norms that different agents can share. Once we have removed the unifying label "norm", we are left with different sets of first-order and second-order dispositions that explain the behaviour of different agents without calling on any sort of shared entity. Therefore, the rule-following paradox fades away, because rules themselves disappear.

There is another notable consequence: the quasi-naturalist reading seems to entail it being possible, at least in principle, to build up something like an abstract set of instructions that enable an *agent* (be it a kind of automaton or a living creature) to *display* rational behaviour. For, if a given disposition is conceived as a function $D: S \rightarrow B$ from states of affairs to behavioural responses, that is, a set of pairs $D \subset S \times B$, where S is a set of states of affairs and B is a set of patterns of behaviour, we can say that an individual x possesses a disposition D only if for each $(s, b) \in D$, if x is in the state of affairs s, she will act in accordance with the pattern of behaviour b.[17] Since the theory of computation tells us that any problem that can be solved by a series of orderly steps can be solved by a Turing machine, the dispositionalist interpretation of attitudes allows us to work out an abstract recipe for building up an agent capable of displaying rational behaviour. The achievement of this task is enough to endow our agent with a complete set of functions $D: S \rightarrow B$ along the lines sketched out in the previous section. The output should be a system capable of mastering the fragment of vocabulary for which it has been taught the relevant semantic content.

Such an upshot does not, *per se*, run counter to Brandom's project.[18] It is even in tune with the abstract setting of Brandom's semantics, which is very akin to the central tenets of classical AI in maintaining that meaning can be brought into existence as a result of the "proper" manipulation of non-semantical items. However, it conflicts with the project, pursued in *BSD*, of under-

---

[17]   That possession of a disposition entails a subjunctive conditional is presumably uncontroversial. I will not enter into the much debated issue of reducibility of dispositions to conditionals, nor is it relevant for the purposes of my argument; for further details, see Mumford 2003; Choi, Fara 2012.

[18]   Remember that Brandom himself, in a recent paper – Brandom 2010: 34 –, has stated that it is possible to arrive at a characterization of sapience in terms of six fundamental semantic and social-pragmatic practices that any kind of agent – be it an extraterrestrial or a digital computer – must display in order to be counted as a potential interlocutor.

standing algorithmic elaboration not on the basis of syntactic computation of symbols, but on the basis of a pragmatic model in which primitive abilities are transformed into more complex ones. In particular, I have in mind the pages of the third lecture in which Brandom attempts a pragmatic revision of some theses of AI functionalism.

Brandom stresses that the practice of adjusting one's other beliefs in response to a change of belief is intrinsically holistic; this raises the problem of revising and updating one's commitments and entitlements in the right way, that is, in a way that is sensitive to one's other collateral commitments and entitlements. Since

any change in *any* property of one changes *some* of the relational properties of *all* the rest ... it is *not* plausible ... that *this* ability can be algorithmically decomposed into abilities exhibitable by non-linguistic creatures (*BSD*: 80-81, italics in text).

Any attempt to deal with this difficulty would have to face the problem of finding a rule to determine what factors are to be ignored. This difficulty is less severe in linguistic creatures; it is the latter that have semantic, cognitive, or practical access to the complex relational properties they would need to distinguish and assess the quality of a number of material inferences (*BSD:* 83).

Brandom contrasts algorithmic decomposition into primitive abilities with training by an expert. A *course of training* may be thought of

as having as its basic unit a stimulus (perhaps provided by the trainer), a response on the part of the trainee, a response by the trainer to that response, and a response to that response by the trainee that involves altering his dispositions to respond to future stimuli (*BSD:* 87).

The abilities involved in this process "vary wildly from case to case, and depend heavily on parochial biological, sociological, historical, psychological and biographical contingencies" (*BSD:* 85). As a consequence, we cannot settle the question of which algorithmic elaboration is sufficient for a particular creature in a particular context empirically.

Brandom's discourse points towards a broader notion of algorithm and algorithmic decomposition, one that can overcome the strictures of the classic conception of algorithms as computable functions.[19] Such a notion is instantiated in the kind of elaboration by training that enables us to develop primitive discursive abilities into more complex ones, such as normative updating. However, the way in which Brandom treats the phenomenon of training suggests that there is nothing magical or mysterious in the manner in which trainers

---

[19] See also Cantù, Testa 2011 on these topics.

instruct their trainees. After all, Brandom himself proposes an abstract model of what a course of training should consist of:[20] a series of responses on the part of the trainee, to which the trainer reacts with appropriate corrections. It seems to me that there is nothing in this succession of events that cannot be algorithmically decomposed. It suffices to set a sequence of stimuli properly arranged so that each stimulus is related to the trainee's previous performance. It is true that the abilities involved vary dramatically from case to case, and that one cannot predict in advance the success of a particular course of training. These empirical limitations do not, however, affect the algorithmic decomposability *in principle* of the training process (the practical implementation of this model can obviously pose almost insurmountable difficulties). They are rather to be viewed as contingent features of the training process. Since a single course of training can be implemented by activating different sets of abilities, the variability of the abilities involved in each particular case is connected, in accordance with the functionalist thesis, with the multiple realisability of the overall process.

In the end, Brandom seems to reproduce a model of explanation that can easily been translated into the dispositionalist jargon. After all, what the trainer is attempting is a appropriate alteration of the (first-order) dispositions of the trainee.[21] It is the practical intractability of the holistic character of our activity of revising our beliefs that motivates Brandom's discomfort with the possibility of an algorithmic decomposition of the ability to engage in an Autonomous Discursive Practice. But whereas it is possible to agree with him that the proper treatment of semantic holism poses a formidable obstacle to our efforts to create a computational system capable of engaging in linguistic practice, no evidence is offered for the stronger claim that such a treatment is *in principle* impossible.[22] The initial plausibility of Brandom's remark that the ability to update one's beliefs appropriately cannot "be algorithmically decomposed into abilities exhibitable by non-linguistic creatures" is based on the implicit contrast between linguistic creatures and non-linguistic ones.[23] But this contrast is misplaced, since we have shown that one of the most important consequences

---

[20]   Brandom's description of training, it may be noted in passing, reminds one of what Donald Davidson, in his later writings, called "*triangulation*". See, for example, Davidson 2001.

[21]   See the quotation above from *BSD*: 87.

[22]   This is, however, a matter for "Utopian Artificial Intelligence", in the sense attributed to it by Stekeler-Weithofer 2008.

[23]   Pirmin Stekeler-Weithofer (2008: 80-81) notes that this distinction is not as sharp as Brandom would have us believe. Automata can be provided with the resources to behave like linguistic creatures. Human infants, on the other hand, are non-linguistic creatures at birth, but quickly develop the cognitive and physical abilities to engage in ADPs.

of Brandom's quasi-naturalism is that it is possible to build up an agent capable of exhibiting rational abilities.

Moreover, the idea that only creatures that have something like semantic access to the complex relational properties they would have to distinguish in order to assess the quality of a number of material inferences are able to engage in ADPs seems to entail an awkward circularity. Indeed, if individuals engaging in linguistic practice must be endowed from the outset with semantic access to the properties they have to deal with in order to acquire new conceptual content, it becomes difficult to see what course of training can create this special insight *ex nihilo.* If the ability to revise and update one's commitments and entitlements is not algorithmically decomposable – in the way the dispositional interpretation seems to suggest – because it involves the capability of identifying, for each new belief, what other beliefs need to be revised, one fails to see what kind of education can teach this ability, since the success of the training process depends on the capacity of the trainee to *revise her beliefs in the light of the suggestions put forward by the trainer.* In response, one might argue that the thesis according to which ADPs are algorithmically decomposable runs the risk of eliminating the *autonomous and the holistic* character of discursive practices. These worries are probably well-founded, but it is not clear from *BSD* what special quality that can prompt autonomous, holistic discursive practice courses of training possess with respect to algorithmic decomposition.

The overall moral of this story can be summarised as follows. I have argued that the failure of Brandom's lines of reasoning about the algorithmic indecomposability of ADPs reinforces a train of thought that a careful reading of his claims against the identification of attitudes and dispositions may suggest. If these arguments fail, and if attitudes are indeed expressible in dispositional terms, and if there are, at the same time, reasons to reject Brandom's scepticism towards the tenets of AI functionalism, then Brandom's pragmatist semantics is intimately bound up with the idea that it is possible, at least in principle, to create an agent capable of exhibiting rational (linguistic) behaviour. The main inspiration of scorekeeping semantics is the idea that we can gain access to meaning through analysis of what we do; this leads directly to the conclusion that sapience is something that can be reproduced, if only we can determine the right moves. What seems to be lacking in this approach, and this absence becomes apparent in the dispositionalist interpretation, is a more direct confrontation with the dialogical dimension of language. Brandom's scorekeeping semantics promises to derive meaning from the social and pragmatic dimension of language. But if my worries are well founded, it stops short of singling out the features that emerge from the polyphonic structure of our linguistic exchanges. Brandom's speakers remain individuals whose linguistic behaviour

may be accounted for in dispositionalist – hence, monological – terms.

4. *Conclusions*

In this paper I have raised a number of doubts on Brandom's pragmatist explanation of norms. In the first section, I briefly reviewed some motivations behind his attempt to locate a middle ground between Platonism and regularism about norms via the thesis of the social institution of normative statuses by the attitudes of the members of the linguistic community. In the second section, I questioned his arguments for the indispensability of normative notions in accounting for normative attitudes. Since the social institution of norms is a matter of possessing the right behavioural first- and second-order dispositions, accepting Brandom's theory means accepting the idea that there may be a story entirely couched in dispositional terms that explains how individuals can institute a whole world of norms, starting from a small set of dispositions to social behaviour and acquiring new dispositions.

I have also maintained that the possibility of giving a dispositional reading of Brandom's semantics is not sufficient *per se* to achieve a complete naturalisation of his pragmatism regarding norms. However, the dispositional interpretation seems to point towards an overall landscape in which linguistic abilities are algorithmically decomposable into primitive skills that do not permit any semantic access. Hence I argued that Brandom's arguments against AI functionalism, offered in *BSD*, are not conclusive. Obviously, the problem of the algorithmic decomposability of linguistic practices cannot be solved within the scope of this short essay. Perhaps it cannot be decided *a priori* either. But there are reasons for considering Brandom's arguments opposing this thesis as unsatisfying.

Ultimately, Brandom's position tends to coincide with a complete reduction of normative attitudes to behavioural dispositions. I do not think that this reduction entails a corresponding naturalisation of normative notions (though dispositions are expressible in non-normative terms) because I am not sure that dispositions are entirely naturalisable. But the primacy of dispositions surely derives from the pragmatist insight that the explanation of norms must evolve from an account of the way in which individuals *use* them.[24] Indeed, commitment to an order of explanation that starts from a description of linguistic practice leads to a picture of normative phenomena centred on individuals' actual behaviour.

---

[24]   Remember the pragmatist thesis that "the *use* of concepts determines their *content*" (Brandom 1999: 164, italics in text).

References

Brandom, Robert B., 1994, *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Harvard University Press, Cambridge MA.

Brandom, Robert B., 1999, "Some Pragmatist Themes in Hegel's Idealism: Negotiation and Administration in Hegel's Account of the Structure and Content of Conceptual Norms", in *European Journal of Philosophy*, 7, 2: 164-189.

Brandom, Robert B., 2008, *Between Saying and Doing: Towards an Analytic Pragmatism*, Oxford University Press, Oxford.

Brandom, Robert B., 2009, *Reason in Philosophy: Animating Ideas*, Harvard University Press, Cambridge MA.

Brandom, Robert B., 2010, "Conceptual Content and Discursive Practice", in *Grazer Philosophische Studien*, 81: 13-35.

Brandom, Robert B., 2011, "The Structure of Desire and Recognition: Self-Consciousness and Self-Constitution", in Heikki Ikäheimo, Arto Laitinen, eds. *Recognition and Social Ontology*, Brill, Leiden: 25-51.

Cantù, Paola, Italo Testa, 2011, "Algorithms and Arguments: The Foundational Role of the ATAI-question", in Frans H. van Eemeren, Bart Garssen, David Godden and Gordon Mitchell, eds. *Proceedings of the Seventh International Conference of the International Society for the Study of Argumentation*, Rozenberg/Sic Sat, Amsterdam: 192-203.

Choi, Sungho, Michael Fara, 2012, "Dispositions", in Edwad N. Zalta, ed. *The Stanford Encyclopaedia of Philosophy*, Spring 2012 edition, accessed 27 January 2014, <http://plato.stanford.edu/archives/spr2012/entries/dispositions/>.

Davidson, Donald, 2001, *Subjective, Intersubjective, Objective*, Oxford University Press, Oxford.

Geach, Peter, 1957, *Mental Acts: Their Content and Their Objects*, Routledge, London.

Grönert, Peter, 2005, "Brandom's Solution to the Objectivity Problem", in *Pragmatics and Cognition*, 13, 1: 161-175.

Hattiangadi, Anandi, 2003, "Making It Implicit: Brandom on Rule Following", in *Philosophy and Phenomenological Research*, 66, 2: 419-431.

Kiesselbach, Matthias, 2012, "Constructing Commitment: Brandom's Pragmatist Take on Rule-Following", in *Philosophical Investigations*, 35, 2: 101-126.

Kripke, Saul, 1982, *Wittgenstein on Rules and Private Language*, Blackwell, Oxford.

McDowell, John H., 1984, "Wittgenstein on Following a Rule", in *Synthèse*, 58, 3: 325-363.

Mumford, Stephen, 2003, *Dispositions*, Oxford University Press, Oxford.

Peregrin, Jaroslav, 2012, "Inferentialism and the Normativity of Meaning", in *Philosophia*, 40, 1: 75-97.

Rosen, Gideon, 1997, "Who Makes the Rules Around Here?", in *Philosophy and Phenomenological Research*, 57, 1: 163-171.

Sellars, Wilfrid, 1954, "Some Reflections on Language Games", in *Philosophy of Science*, 21, 3: 204-228.

Sellars, Wilfrid, 1974, "Meaning as Functional Classification", in *Synthèse*, 27, 3-4: 417-437.

Stekeler-Weithofer, Pirmin, 2008, "The Computational Theory of Mind and the Decomposition of Actions", in *Philosophical Topics*, 36, 2: 63-86.

Testa, Italo, 2011, "Social Space and the Ontology of Recognition", in Heikki Ikäheimo, Arto Laitinen, eds. *Recognition and Social Ontology*, Brill, Leiden: 287-309.

Wanderer, Jeremy, 2008, *Robert Brandom*, Acumen, Stocksfield.

Wright, Crispin, 1980, *Wittgenstein on the Foundations of Mathematics*, Duckworth, London.

Leonardo Marchettoni
leonardo.marchettoni@unipr.it
Department of Law
University of Parma