philosophies

MDPI

Article

Backpropagation of Spirit: Hegelian Recollection and Human-A.I. Abductive Communities

Rocco Gangle

Center for Diagrammatic and Computational Philosophy, Endicott College, Beverly, MA 01915, USA; rgangle@endicott.edu

Abstract: This article examines types of abductive inference in Hegelian philosophy and machine learning from a formal comparative perspective and argues that Robert Brandom's recent reconstruction of the logic of recollection in Hegel's *Phenomenology of Spirit* may be fruitful for anticipating modes of collaborative abductive inference in human/A.I. interactions. Firstly, the argument consists of showing how Brandom's reading of Hegelian recollection may be understood as a specific type of abductive inference, one in which the past interpretive failures and errors of a community are explained hypothetically by way of the construction of a narrative that rehabilitates those very errors as means for the ongoing successful development of the community, as in Brandom's privileged jurisprudential example of Anglo-American case law. Next, this Hegelian abductive dynamic is contrasted with the error-reducing backpropagation algorithms characterizing many current versions of machine learning, which can be understood to perform abductions in a certain sense for various problems but not (yet) in the full self-constituting communitarian mode of creative recollection canvassed by Brandom. Finally, it is shown how the two modes of "error correction" may possibly coordinate successfully on certain types of abductive inference problems that are neither fully recollective in the Hegelian sense nor algorithmically optimizable.

Keywords: abductive inference; machine learning; G.W.F. Hegel; Robert Brandom; human-A.I. interaction

check for updates

Citation: Gangle, R. Backpropagation of Spirit: Hegelian Recollection and Human-A.I. Abductive Communities. *Philosophies* **2022**, *7*, 36. https://doi.org/10.3390/philosophies7020036

Academic Editor: Woosuk Park

Received: 31 January 2022
Accepted: 24 March 2022
Published: 26 March 2022

1. Introduction

Communicative interactions between human agents and interfaces driven by machine learning algorithms are becoming commonplace in contemporary global society, and it is almost certain that the prevalence of such interactions only stands to increase in the near future. The present discussion is concerned with the dynamics of abductive inferences made in these sorts of human–machine interactive contexts, particularly when human agents and machine learning processors might to some degree work collaboratively in generating abductive inferences. It will be argued that the cognitive framework of *recollection* that Robert Brandom reconstructs in [1] from Hegel's *Phenomenology of Spirit* [2] provides a way to specify one mode of potential interaction between human and machine learning agents and within that specific context to delimit certain possibilities and obstructions. It is important to emphasize at the outset that within the limits of a single article it will be impossible to deal in a comprehensive way with Brandom's complex analytic-pragmatic reading of Hegel, much less the enormous body of scholarly literature that has been built up around the *Phenomenology* itself. Neither will it be possible to provide a thorough analysis of abductive inference in general, although it will be feasible to specify a set of constraints determining what will count as an abduction for the purposes at hand. What is at stake in the present argument is to isolate, from among the many potential types of human–A.I. collaborative interactions that might involve abductive inferences, a single definite kind of abductive reasoning process and to show how the roles of responsibility and authority are distributed within it based on its intrinsic features. The kind of collaborative abductive reasoning that will be specified is that of communities of agents that may be partitioned into sub-communities of *recollective* and *non-recollective* agents. It will be shown that for such communities and collaborative processes, the sub-communities composed of recollective agents must, for reasons that will elaborated on, be the ones who possess responsibility and authority for crucial aspects of the overall process simply by virtue of those sub-communities being recollective in form.

The presentation proceeds as follows. The first section summarizes Brandom's reconstructive reading of Hegelian recollection and recasts that reading in the context of abductive inference. Recollection is recast here as a specific type of abductive inference, which will be called *recollective abduction*. Brandom's own privileged example of Anglo-American case law is used to illustrate the ideas and structures at play and to make the notion somewhat more concrete, as well as to introduce the notion of an *institution-frame*, or *i-frame*, as a way to tease out an important difference between external and internal constraints on the dynamics of recollective reasoning. The point of this initial section is to show that Brandom's analytic-pragmatic reconstruction of Hegelian recollection may be understood as a particular type of abduction and that the type of abduction thus specified may involve its own peculiar dynamics when incorporated into a community consisting of both human agents and A.I. (machine learning) agents or processes. The fact that recollective abduction involves an explicit narrativization of the deliberative process that produces it introduces particular challenges for including algorithmic agents in the communities who make such abductions. Nonetheless, the idea at the heart of recollection that cognitive progress can be made by means of iterative processes of error-correction that revise previous stages in a multi-stage cognitive operation suggests an analogy with the machine learning technique of backpropagation. As in the machine learning technique of backpropagation, outputs are generated in recollective abduction by treating a multi-stage operation as subject to revision stage by stage according to the controlled recalibration of cumulative effects. When the inputs and outputs of backpropagation processes are composed with one another and with the premises and conclusions of human reasoning, who or what holds the cards and with what justification? The distinction between recollective reasoning and backpropagation learning must be carefully delineated in order to answer this question properly.

Accordingly, in the subsequent section, the structure of recollective abduction drawn from Brandom's work is contrasted with the algorithmic method of backpropagation as commonly used in machine learning neural networks. Despite certain formal similarities between the two processes, it is shown how the crucial ingredient of the self-reflexive narrativization of the learning process is inherently missing from the implementation of backpropagating machine learners, whatever their other epistemic and even potentially abductive virtues, thus meriting the designation of their type of reasoning as *non-recollective learning*. In this way, the contrast between the kinds of human agents and communities that perform abductive inferences in the form of recollection and A.I. learners that develop cognitive competencies by way of backpropagation methods is made precise by identifying exactly what it is that distinguishes them. This mark of distinction is not some mysterious property such as consciousness but is rather an objectively determinable feature of the cognitive processes themselves.

Finally, then, given the distinction between recollective abduction and non-recollective learning, the problem may be posed of how communities of interacting agents consisting of both types of reasoning might possibly collaborate successfully with respect to common problems requiring the kinds of abductive inferences such communities are likely to make. In other words, it may be asked: How do processes of recollective abduction made in the context of human–A.I. collaborative interaction necessarily work? Are there intrinsic limits or constraints in such contexts? The main finding of the paper is that under such conditions the intrinsic asymmetry between recollective and non-recollective subcommunities generates dynamics of evaluation, inclusion and exclusion that can only be moderated and decided on by the recollective subcommunities themselves. In short, the very structure of recollective abduction excludes non-recollective learners from participating on an equal footing with recollective agents. Communities of human–A.I. interaction of this type are certainly possible, and machine learning agents might contribute significantly to various aspects of the collaborative cognitive tasks involved. Nonetheless, the self-narrativizing role of recollective agents gives them (and the sub-communities they form) a unique and unimpeachable authoritative status for any such human–A.I. collaborative community. So long as machine learners remain non-recollective, which is an objectively specifiable characteristic, not a vague or indeterminate property such as self-awareness, they will be capable of contributing to collaborative human-A.I. (machine learning) communities only within the limits thus discerned. A series of brief remarks concerning the potential role of the creation and codification of new *i-frames* in managing such situations suggests possible directions for further research and concludes the paper.

2. Recollective Abduction: Brandom's Reading of Hegel

In [1], Brandom recasts Hegel's dialectic of spirit in the idiom of Brandom's own broader project of analytic pragmatism. This reconstruction of Hegel builds upon Brandom's early theorization of the integration of normative pragmatics and inferential semantics in *Making it Explicit* [3] and also deepens the series of reconstructive readings from the history of philosophy that make up *Tales of the Mighty Dead* [4]. Indeed, *A Spirit of Trust* [1] may be understood as a synthesis of the (mostly) formal analyses of [3] and the (mostly) interpretative historical studies of [4]. Moreover, this synthesis plays an important role in retrospectively justifying both sides of Brandom's earlier work. In a certain sense, the reading of Hegel's *Phenomenology* serves to explain both how and why Brandom's earlier conception of analytic pragmatism became possible as a determinate program in contemporary philosophy as well as why and how that program's culmination in a reconstructive reading of Hegel in particular became necessary. Brandom's program of analytic pragmatism extends the work of his teacher and colleague Wilfrid Sellars, which is itself grounded in a combination of the analytic philosophy of language (strongly influenced by [5,6]) on the one hand and Peircean pragmatism on the other (as detailed in [7]). The roots of Brandom's understanding of Hegel are thus to be found largely in philosophical traditions not typically associated with Hegel and German idealism. These links between Brandom's distinctive reading of Hegel's *Phenomenology* and his broader program of analytic pragmatism help to distinguish Brandom's approach from other recent Anglophone Hegelian philosophers such as [8,9]. It also provides an implicit connection to Peirce's work on abductive inference, which supports the interpretation of recollection developed here in terms of abduction.

This movement of retrospective and seemingly anachronistic explanation that becomes at the same time a mediated *self*-explanation is at the heart of Brandom's analysis of Hegel. It is characteristic not only of what Brandom *says* is the best way to conceive of Hegel's project; it is also what Brandom claims to be *doing* with respect to the Hegelian philosophical corpus. In this way, Brandom does not only offer an analytic pragmatic reading of Hegel. Rather, in doing so, he at the same time and by the same means intends to demonstrate how the formally articulable process of pragmatic and historically mediated revision of inherited concepts that he identifies as being central to Hegel's thought is equally at work in his own philosophical project of creative reconstruction as applied to Hegel's thought. Brandom's analytic pragmatic interpretation of Hegel is thus just as much an Hegelian reading of analytic pragmatism.

How is the interpretation structured? Brandom himself notes that his reading of Hegel is organized around a trio of "master ideas" drawn from Hegel's own library of concepts but reworked by Brandom himself into a new synthesis. Those ideas include *determinate negation* as the basis for a conceptual semantics, *mutual recognition* as the root of normativity at the heart of social pragmatics, and *expressive recollection* as the narrativizing process that coordinates the semantic and pragmatic dimensions of the account by way of historical institutions [1] (p. 636). In the present context, we will attend essentially only to the third of these three notions. It is by Brandom's own reckoning a kind of "keystone of the edifice" [1] (p. 637), and in addition to surveying this core concept, there will be space to make little more than passing reference to mutual recognition as well.

The core of expressive recollection is the idea that all epistemic judgements that have, as a matter of fact, been made possible by some experiential, deliberative or self-corrective process of whatever kind *must,* if they are to be appropriately justified, be underwritten by an explanatory account that serves to legitimate that process retrospectively by rendering it coherent in light of a narrative of progressive experimental discovery, that is, *as a story of learning*. As Brandom puts it:

> "To be entitled to claim that things are as one now takes them to be, one must show how one *found out* that they are so. Doing that involves explaining what one's earlier views got right, what they got wrong, and why. It involves rationally reconstructing the sequence of one's previous views of what one now takes to be the same topic so as to exhibit it as a process of *learning*, of gradual *discovery* of how things actually are." [1] (p. 680)

Brandom's viewpoint of pragmatism leads him to conceive of epistemic questions in terms of practical commitments. Thus, given an experience of error, that is, the finding of oneself in the uncomfortable pragmatic situation of simultaneous commitments to mutually incompatible courses of action (corresponding in general to incompatible epistemic views), one must typically make a choice to hold to and act upon only certain commitments and to give up one or more others. In the wake of such an adjustment of commitment to relieve the incompatibility, it becomes necessary on Brandom's Hegelian account to construct an explanatory narrative that makes sense of the adjustment as not only epistemically plausible (more likely to be in accord with how things really are) but, more importantly, epistemically enriching (that is, more clearly illuminated now by virtue of the experiential confrontation with the incompatibility and its subsequent overcoming via the specific choice made). In other words, on Brandom's account, it is insufficient merely to accommodate incompatible commitments by choosing among them or reforming them in some way so as to resolve the incompatibility, even if such adjustment is in some relevant sense perfectly correct. It is necessary in addition to convert this mere transition from one state of commitment to another into the story of a *process of learning* by providing a narrative according to which the prior state of incompatibility has become, through this very process, an opportunity for understanding the relevant conceptual contents in a new and better way.

To make Brandom's main idea here more vivid, it may be useful to consider a simple example. A father and son are watching fish swimming in a stream. The son tries to catch a fish with his hands but finds that he consistently misses his target. The father suggests aiming several inches *deeper* than the fish appears to be. Following his father's advice, the son successfully captures a fish with this new method. However, the new method is not fully understood until the father *explains* that and how the previous difficulty (repeatedly missing the target fish) was the result of the refraction of light in water. By providing not only a successful new method but also an explanatory narrative according to which the new method may be understood explicitly as correcting an error implicit in the original method, the father and son together have *made progress* both practically and epistemically.

Such an explanatory conversion of conceptual adjustment into epistemic progress is what Hegel calls *recollection* (German: *Erinnerung*). It is what, according to Brandom, "turns a *past* into a *history*" by generating a narrative as a "narrative of *expressive* progress" [1] (p. 681). What this means is that epistemic progress in the mode of recollection is not merely the substitution of a better (more accurate, more coherent, etc.) epistemic state for some prior worse state. Recollection adds to such a transition from a worse to a better state of knowledge an account of *why* and *how* epistemic progress has been made by means of error and its overcoming. This explanatory dimension takes the form of a subjectively determining narrative that partly defines the epistemic agent as a subject who has learned as well as an objective account of what features of the object of knowledge have been made more explicit and how. What makes this type of narrative *expressive* on

Brandom's account is that it is constrained to relate earlier to later states of knowledge in terms of an underlying *continuity* of conceptual content. Error is overcome with respect to *something* about which one had previously been in error and which one now understands more clearly. In other words, the narrative must tell the story of how one and the same conceptual content has come to be better expressed by the later understanding. This expressive account equally involves what is known and the knower who knows it. Its two sides are in this manner the clearer or more accurate expression of some determinate conceptual content on the one hand (the objective side) and the enrichment of an ongoing narrative of the learning self (the subjective side) on the other.

The first core claim of the present argument is that the process of Hegelian recollection as analyzed by Brandom may be understood as a particular type of abductive inference. Given the complexity of the notion of abduction and the lack of a single, generally agreed-upon theoretical paradigm for modeling abductive inference, it is necessary to approach the problem from a bird's-eye view. We proceed to establish the relation between abduction in general and recollective abduction in particular as follows. First, we provide a schematic account of how abduction in general may be understood for the purposes of the present argument. Next, we show how Hegelian recollection in the sense developed by Brandom and outlined above may be conceived as a specific instance of that schema that adjoins additional determinate structure to the more general schematic account. Finally, a particular case used by Brandom himself shows, concretely, in what sense the recollective process is susceptible to description as a specific type of abductive inference involving well-determined additional features.

2.1. A Schema for Abductive Inference

What is abductive inference? The notion of abduction as a third mode of inference to be thought of systematically alongside deduction and induction may be traced back to C.S. Peirce, although the general problem of hypothesis formation as a key ingredient of reasoning traces back to much earlier in the history of philosophy. In a well-known formulation of Peirce's [10] (Vol. 2, p. 231), the explanatory character of abduction is best characterized as a cognitive response to a surprising fact:

The surprising fact, C, is observed;
However, if A were true, C would be a matter of course.
Hence, there is reason to suspect that A is true.

This picture of abduction is a useful starting point for grasping the notion, but it by no means clarifies exactly what abduction's scope and limits actually are. Indeed, the question of what abductive inference actually consists of remains largely open. There have been various attempts to characterize abductive inference in the context of formal logic, such as [11]. Much work in recent decades has been devoted to developing more thorough theories of abduction, such as [12–14]. Especially helpful is [15], which provides a sketch of a taxonomy of different types of abduction. Recently, the papers collected in [16] treat a variety of issues coordinating epistemic and pragmatic concerns with respect to abductive inference.

For the purposes of the present paper, it will be convenient from a bird's-eye point of view merely to specify a schema of abductive inference consisting of four phases:

1. **Occasion**: Abduction begins with the introduction of new information into the purview of some cognitive agent. This starting point corresponds to Peirce's notion of "surprising fact," although it is important to note that the affective response of surprise is not yet registered in this phase. The surprise as occasion is understood as an event and not yet an affect. To use a term from Sellars' theory of inference [17], such an occasion serves as a "language-entry" event.
2. **Discrepancy**: The mere introduction of new information is not sufficient to warrant an abductive inference. The provocation for abduction consists of a discrepancy that is registered by the relevant cognitive agent between the newly introduced information and some measure of "expected" value. It is in this phase that the subjective affect

of surprise may occur, although from the present standpoint such an affect is not a necessary component.

3. **Conjecture**: This phase is the core of abductive inference. Here, a hypothesis H is formulated such that it would, if true, resolve the discrepancy noted in the second phase. Typically, a requisite constraint on H is that it is structured so as to resolve the discrepancy between the new and the previously held or expected information by *explaining* the new information in some relevant sense, that is, by linking it appropriately to information concerning other facts or principles that are considered reliable by the agent. Such an explanation may involve retracting or modifying previously held beliefs of the agent.
4. **Discharge**: The pragmatic role played the conjectured hypothesis H is a source of considerable controversy in the debates over abductive inference. Here, it is presumed that H must eventuate in some pragmatic *program* that guides future action in some respect relevantly controlled or guided by H. Typically, such a program would consist of some form of inquiry that would aim to establish or to annul the epistemic reliability of H. In a way that is complementary to the "language-entry" character of the initial Occasion, the Discharge serves, in Sellars' terminology [17], as a "language-exit" event.

This schema is not intended to be applicable to every possible modality of abductive inference. It is unlikely that any such schematization of abduction is even possible. Nonetheless, this particular viewpoint on abductive inference is taken here because it possesses at least two virtues. First of all, it draws attention explicitly to the role of abduction as a kind of "inferential subroutine" in the logical space of reasons with (passive) inputs from and (active) outputs into the physical environment and its space of causes. This is why reference was made to the work of Sellars in the schema above. Secondly, distinguishing between the two phases of Discrepancy and Conjecture mitigates any notion of abductive inference as a purely ad hoc explanatory invention. Rather, since abductive conjecture is here motivated by a specific discrepancy in epistemic values or expectations, the content of the conjecture will naturally be understood as Janus-faced in the sense that it is directed on the one hand to explaining the surprising fact (as in more traditional accounts) but also oriented on the other hand to the fine structure of the epistemic context in which the surprising fact was, in fact, surprising in the first place. In this regard, abductive inference requires attention to the form of its *explanans* as well as the content of its *explanandum*. Erstwhile conceptions, even poorly justified ones, are seldom given up easily by cognitive agents, and there is a conservative epistemic inertia in the logical space of reasons somewhat analogous to the inertia of momentum in the physical space of cause and effect. This conservative tendency constrains the field of what might qualify as a viable conjecture H, in the present account. Roughly speaking, the epistemic adjustment that would be required to assimilate H as, in fact, true should be more or less commensurate with the magnitude of the discrepancy registered in the second phase. These notions can be made more precise, as they are, for instance, in [12–14]. Those details are not especially relevant here, however, since what is at stake is not an attempt to evaluate abductions (for example, to aim to analyze what "best" might mean in the context of "inference to the best explanation") but only to characterize certain essential features of abductive inference. The features that are relevant to the present account are precisely those that are named by the four phases outlined above. Whatever else might be important and indeed crucial for analyzing abduction as such, the claim at stake here, is that these four phases that are recognizable features of at least a large and representative class of abductive inferences should remain relatively uncontroversial. The purpose of this claim is not to make a contribution to the literature on abduction in general, but to provide general parameters for understanding the case at hand, namely a specific type of human–A.I. (machine learning) collaborative task coordination.

2.2. Recollection as a Mode of Abduction

Given the framework outlined above, it is possible to reframe the earlier account of Hegelian recollection in terms of the structure of abduction, indeed as a specific type of abductive inference. We will call such a mode of abductive inference a *recollective abduction*. What determines such a mode of abductive inference? Broadly speaking, what is added to the four phases described above is an emphasis on the fact that the recollective abductive process takes place in a community of inquiry, and thus involves a definite social, communicative and deliberative context; and that because of this, the formulation of the conjectural hypothesis *H*, phase 3, must be made publicly in that social context. The public and social character of the recollective abductive process will entail that the self-definition of the community as such is at least partly at stake in the process. In this sense, recollective abduction must be understood as a constitutive component of the ongoing interpretative process of a *tradition*.

First, however, it is necessary to chart more precisely how the Hegelian process of recollection fits into the schema of abductive inference described above. It is clear that the initial phases of Occasion and Discrepancy accommodate Brandom's reformulation of Hegel quite well. In particular, the roles of occasion and discrepancy are specified and made somewhat precise in Brandom's account by means of the definite notion of the *material incompatibility* of commitments. A Discrepancy is evident in a given epistemic/pragmatic context when two or more commitments are affirmed by an agent (or some appropriately integrated community of agents) and the set of what is necessarily implied by one of the commitments intersects with the set of what is incompatible with (that is, necessarily excluded by) one or more of the others. For example, a commitment to the claim "this paper is flammable" is incompatible with commitment to the claim "if I bring this paper into contact with the candle flame, it will remain unburned." A richer, detailed formalization of this idea of material incompatibility is provided in Brandom, *Between Saying and Doing* [18]. Furthermore, a reconstruction of the epistemic dynamics that result from this key idea may be found in [19]. Given the formal structure of the incompatibility of commitments as a starting-point for integrating Brandom's account into the framework of abduction, what is most important for conveying what is specific to recollection as such is its communal deliberative and expressive character. In any case, what should be clear is that Brandom's notion of the material incompatibility of commitments may be understood to function as both Occasion and Discrepancy.

The phases of Conjecture and Discharge are also recognizable in Brandom's analysis of Hegel's notion of recollection to the extent that in the face of whatever situation of the material incompatibility of commitments, on Brandom's account, a decision must made be made by the relevant agent or community of agents that (1) resolves the incompatibility by choosing among the commitments and (2) justifies that choice explicitly by way of a "Whiggish" narrative that makes the choice appear epistemically progressive. The role of the hypothesis *H* in the general schema is played here by the determinate choice of whatever commitment overrides its incompatible rival. That choice becomes explanatory to the extent that the narrative accompanying it is successful at justifying its epistemic advantages.

To specify in detail, then, *recollective abduction* should be understood as a form of abductive inference that adjoins to the four schematic phases as described above with the following additional criteria:

1. The present sequence of Occasion, Discrepancy, Conjecture and Discharge is embedded in a *series* of such abductive processes that are linked one to the next so as to constitute a *unified tradition* with its own internal structures, conceptual parameters and protocols. Part of what unifies this tradition is the fact that each discrepancy in the series is determined with regard to an earlier result or consequence internal to the tradition itself. In this sense, the tradition is intrinsically self-reflexive and self-correcting.
2. The Conjecture phase includes an explanatory account that not only works potentially to justify whatever position is taken with respect to the given discrepancy (and thus already, in part at least, begins to resolve the cognitive dissonance resulting from the discrepancy), but one that also establishes how the current inference in the series maintains a conceptual continuity with earlier inferences in the series, both in spite of

and yet also, by the same token, in virtue of their now-discerned errors. In this way, recollective abduction introduces an additional constraint on the form of conjecture necessitated by the abductive process as such. This additional constraint involves the articulation of a *narrative* that, if true, would include the current conjecture H in an ongoing progressive or "Whiggish" account of conceptual understanding by the community as a whole that would both extend and retrospectively clarify the earlier stages of its self-constitution as a tradition.

3. The Discharge phase includes an explicit presentation of the narrative of progressive change on the basis of traditional continuity in a way that is appropriate to the type of tradition and the kind of community that is affected by the conjecture. In this respect, whatever other features the discharge of the abductive hypothesis might have, when made in the context of recollective abduction, the final phase must also explain and thereby justify *itself* as a conjectural program worthy of the community's commitment in light of its shared tradition.

Crucial to the Hegelian account of rational spirit and in particular the structure of what we are calling here recollective abduction is the social character of conceptually articulated experience. Recollective abduction takes place in a community of agents and occurs by way of communication and coordination among those agents in the context of a determinate tradition. In this respect, recollective abduction is not a psychological act but a social practice. In particular, it depends crucially upon the Hegelian dynamics of *mutual recognition* that Brandom examines and reconstructs in a more analytic idiom in chapters 8–12 of [1]. Brandom himself provides a complex formalization of this dynamic of social recognition via complementary attitudes and statuses. Foregoing a detailed examination of this aspect of Brandom's interpretation, the irreducibly social character of recollective abduction can be best introduced here and readily understood by way of a concrete example.

2.3. Anglo-American Case Law as an Instance of Recollective Abduction

Brandom returns to the example of Anglo-American case law as illustrative of Hegelian recollection several times throughout [1]. Indeed, this example takes on a sort of canonical status in Brandom's account. The dynamics of case law show, in a particularly perspicuous way, how Hegelian recollection in general is understood by Brandom to function, and the structure of mutual recognition, especially in its historical dimension, is made directly evident in the relations that the juridical process gives rise to among the members of the legal community. It may be remarked that a more detailed study of this particular example is provided in [20].

In Anglo-American case law, legal judgments are not understood as punctual applications of a general law (for instance, a law prohibiting theft) to particular cases (such as, say, the specific event of Ringo stealing John's guitar). Instead, each legal judgment refers explicitly to earlier judgments that are deemed relevantly similar to it. A judge in such a system of legal reasoning must articulate a decision that explains how the verdict with respect to the case at hand has been derived from the verdicts (and explanations) of previous cases, which thereby take on the role of *precedents*. There is thus a nested series of retrospective references that, taken together, constitute a juridical tradition. Since the derivation from the precedent of the verdict in each case is creative and interpretative, the reasoning involved is essentially abductive. Each juridical decision reinterprets the previous verdicts at issue as stages in a process that leads, under this very interpretation, to the explanatorily elaborated correctness of the present verdict. In particular, then, the juridical community constituted by such a series of abductive judicial decisions may be seen to be recollective in the sense of Hegel and Brandom outlined above. In particular, Brandom's choice of case law as a privileged example of Hegelian recollection may be contrasted with Hegel's own theory of law as presented in the *Philosophy of Right* [21]. Although Hegel does understand law as synthesizing both formal and cultural/material aspects, his own constitutionally based model is less self-revisionary and hence less abductive than Brandom's Anglo-American model.

Nonetheless, similarly to Hegel's own account of state legal authority, the protocols of Anglo-American case law are in a way intrinsically conservative, in the sense that any radical or complete break from past tradition is excluded a priori. This essential conservatism is strongly linked to the way case law as an institutional framework of recollective abduction at once depends upon and determines a definite sub-community of jurists within the larger community of some given society or nation-state. Every particular case decision must, by virtue of the very form of reasoning that constitutes the system of case law as such, justify itself before the community of jurists (both present and future) as a partial self-description of the ongoing process that is both condition for and expression of the pragmatic role of that very community for the surrounding society at large. Such a justification takes the form of an interpretative argument that takes earlier decisions in the tradition as premises (as legal precedents) and explains how those earlier decisions should be understood to lead, under the appropriate redescription of the tradition's ongoing self-definition, to the current judge's own decision. In this way, every decision ideally affirms itself as the proper present decision for the tradition by offering reasons to the community who constitute that tradition as to why that community should understand itself as having been correctly redescribed in precisely the terms necessary to accept those reasons. In a sense, the reasoning involved here is circular. However, it is circular only because the community for whom the reasoning is validated is also the community constituted and maintained by that very act of validation.

Two remarks are in order with respect to this privileged example of case law. First of all, it should be noted that the dynamics of recollection are supported in the instance of case law by the institutional structure that locates a particular functional role for this process in the context of a broader social context. In other words, recollection works the way it does in this instance because it operates within a clearly defined social "subroutine" with determinate inputs and outputs. Whatever the complex details of any given case may be, its status *as a case* is determined by its inclusion within a juridical institutional structure with well-defined rules and procedures. Similarly, once a judgement is made by a judge, the status of the judgement *as a legal judgement* depends entirely on the same structure.

The second remark follows upon the first. The institutional shell of juridical structures, certifications, procedures and legitimacies provides a distinction between communicative acts and operations that occur strictly within that context and those that do not. A judge, for instance, might ask the bailiff to bring her a glass of water during a trial, but such a communicative act does not, strictly speaking, occur as part of the proceedings. Thus, the institutional-dependence of recollection entails a distinction within the background condition of mutual recognition. Mutual recognition does not simply hold within a community of agents as a universal "flat" relation between all pairs of agents (as Brandom himself unfortunately seems at times to suggest). Rather, institutions themselves generate internal communities with specific roles that require mutual recognition in their own right (such as judge, bailiff, lawyer, defendant), while the existence and proper functioning of those very institutions requires legitimation by means of a different (typically more general) community of mutual recognition for whom the institutions themselves are recognized and their roles and proper limits designated.

It seems that Brandom's choice of this particular example serves thus to emphasize a certain dependence relation between recollective processes and social institutions. If this dependence is a mere accident of the example, then the example itself would risk begging the question of whether this special feature is in fact essential to the process. Instead, by a more generous reading of Brandom, the example may be understood to make especially salient a feature of all such processes, namely their embeddedness within relatively fixed social institutions. On this reading, institutional frames in the sense of relatively fixed protocols for sequences of action and roles of authoritative decisions and declarations are an essential component of recollective abductive communities insofar as the latter are necessarily tradition-bound. In order to correct this important notion conceptually, we will introduce the term *institutional-frame*, or *i-frame* to describe such structures. An i-frame serves as a more or less fixed shell that compartmentalizes practices, decisions and communicative acts along distinct channels for a given sub-community's delegated pragmatic functions for a larger community. For a similar perspective on social and business organizational structure as coordinated with the basic operation of decision-making, see the analysis in [22]. The ideal type of an i-frame would take the form of an algorithmically decidable flowchart distributing complex tasks through well-defined sequences of sub-tasks, each with its definite requirements for input and output. This compartmentalization of social protocols might suggest some connection with formal cybernetics and the algorithmic operations of computation.

3. Backpropagation Algorithms and Recollective Abduction

With the framework of recollective abduction in place, the argument turns now to the dynamics of machine learning and in particular the well-known method of backpropagation in multilayer neural networks. We will examine the technique of backpropagation in light of Brandom's reading of Hegel, as discussed above. The emphasis on the particular concept of Hegelian recollection differentiates the present argument from more general applications of Hegel's thought to the problems and possibilities of artificial intelligence, such as those in [23,24] (which critically evaluate the proposals in [25,26]).

The technical details of how backpropagation algorithms are implemented in neural networks are the scope of the present paper. Detailed introductory accounts may be found in [27,28]. A helpful survey of the mathematical tools requisite for the implementation of machine learning systems is available in [29]. In order to bring these algorithms into the purview of the present account, the following brief summary suffices. What matters for this account is the manner in which a relatively coarse-grained description of the process may itself suggest how the progressively error-correcting processing of such neural networks might be more or less naturally compared with and partly subsumed under the previous description of Hegelian recollection. Arguably, the typical tasks for which machine learning systems consisting of backpropagation algorithms are trained fit the schema of inductive reasoning rather more closely than that of abductive reasoning. Training an A.I. machine learner to distinguish images of cats from those of dogs, for instance, seems to be a cognitive task more closely aligned with generalization than with explanation. Nonetheless, any such implementation of an algorithmic process to fulfill such a task may be understood as being abductive in principle to the degree that the trained network is intended to function successfully with regard to new data that is sufficiently dissimilar to its original training data. The trained network as a whole may in this respect be understood as a type of abductive hypothesis with respect to the successful fulfillment of relevantly similar tasks. Of course, the network itself does not understand itself in this way, but external trainers and collaborators might very well see things in such a light. We will examine the limits of such a conception in more detail below.

The following account will (1) summarize the general features of multilayer neural networks; and (2) describe the training method of backpropagation informally as an iterated two-stage algorithmic procedure.

A multilayer neural network consists of a sequence L_i of layers, where the index i ranges from 0 to n. L_0 is called the input layer, L_1 to L_{n-1} are called hidden layers, and L_n is the output layer. Each layer L_i consists of a fixed number of neurons, each of which may be understood as a variable function taking values that are output by neurons in the previous layer L_{i-1} as arguments and generating a real number r as value that serves in turn as one of the argument inputs for one or more (often all) of the neurons in the subsequent layer L_{i+1}. The input and output layers are exceptions: the neurons of the input layer are assigned values directly by an external user (the supervisor); whereas the neurons of the output layer are not input into any further function and are read off as, precisely, outputs of the entire process. In this way, the entire sequence of layers may be understood as constituting a single function taking a vector of inputs in the input layer as argument and generating a vector of outputs in the output layer as value. This function is distributed throughout the hidden layers in such a way that the global function $L_0 \rightarrow L_n$ is in fact calculated through a series of intermediate local functions $L_0 \rightarrow L_1 \rightarrow L_2 \rightarrow \ldots\ L_{n-1} \rightarrow L_n$. The global function is thus simply the composition of the sequence of local functions. At each layer, the function that sends values to the next layer is determined by a set of *weights* associating each neuron with, very roughly speaking, a coefficient that scales up or down the relative importance of that neuron's output value as an input for the following layer.

The method of backpropagation uses an algorithmic procedure to systematically alter the weights of the neurons in the network in order to progress step by step towards an adequate approximation to some given target global function $L_0 \rightarrow L_n$ of inputs at the input layer to outputs at the output layer. This target function serves as a set of training data, and the process of iterating the backpropagation algorithm so as to approximate this function constitutes the training of the network. More complete details of how backpropagation is realized mathematically may be found in [27] (pp. 185–230) and [29] (pp. 138–143). Here, the process is schematized in a somewhat rough and informal fashion, solely in order to suggest its similarities with the process of recollection. We describe the algorithm at two levels, external and internal.

At the external level, the algorithm essentially consists of three steps that are repeated as many times as necessary to achieve an approximation that is deemed sufficiently close to the target function for whatever purposes. It is assumed that there is an initial distribution W of neuron weights, typically generated at random, and that there is a fixed input vector I at the input layer that serves as training data coordinated with a specific target output vector T.

Step 1: Generate an output vector V by running the network using the input vector I and the current weights W.

Step 2: Calculate the error E by comparing V with the target vector T.

Step 3: Recalibrate the weights W in such a way that the error E will be reduced.

The sequence of steps 1 through 3 is repeated as many times as necessary to approximate the target function $I \rightarrow T$. Since the error E is reduced on each pass, in principle, the error can be minimized to any requisite margin.

The backpropagation method itself provides an algorithmic process for accomplishing this three-step external procedure, particularly for recalibrating the neuron weights W in step 3 in order to reduce the error E. We conceive of this algorithm as functioning internally in roughly the same sense that a car engine is the internal process that (in this case, literally) drives the external behavior of a car as getting from place to place. At this more detailed internal level, the backpropagation algorithm effectively occurs in two stages, namely by a "forward pass" followed by a "backward pass":

Forward pass: This stage corresponds essentially to step 1 outlined above. The values at the input layer (that is, the arguments of the network function) are input as arguments to the function defined by layer 1; the values of the function at layer 1 are input as arguments to the function at layer 2; and so on. Importantly, not only do the weights in W determine each neuron, which is recorded for later calculation, but also the local value of each function at every neuron, which is stored in memory in order to evaluate each neuron's specific contribution to the global function.

Backward pass: Once the output values and resulting error have been calculated, the recalibration of weights proceeds from the last hidden layer "back" to the first. It is this reversal of the direction of error response and recalibration as compared to the processing of the network itself that suggests the name "backpropagation". Beginning with layer L_{n-1}, the weights of the neurons are adjusted in such a way that, given the new values to the weights, the error E of the output vector V is reduced. In other words, when the output calculated with the new adjusted weights is compared with the output generated by the previous weights, the new error, which may be designated E', is smaller than the previous error E. With the new weights for layer L_{n-1} thus determined, those newly adjusted weights are then "propagated back" to layer L_{n-1}. The weights for the neurons in layer L_{n-1} are then adjusted in such a way that they accommodate the newly adjusted weight-values in layer L_{n-1} and, taking those new values into account, further reduce the error value from E' to some smaller value E''. This process continues in the same manner layer by layer until the adjustment of the weight-values of the neurons in layer L_1 is finally achieved. With all the weights of all the layers now appropriately adjusted, the network is updated and is prepared for a new iteration of the process starting from a new "forward pass" through the updated network.

It is important to acknowledge that the functions determining each neuron in every hidden layer are subject to revision so as to approximate the externally given function $I \rightarrow T$ by means of which the network as a whole is "trained". Intuitively, the values of this training function are taken to be the "correct" values in whatever context of learning, and the fine-tuning of the functions at each neuron in the hidden layers aims to adjust the distribution of the computational task across the entire network in a way that is sufficiently robust to accommodate new data sets that are relevantly similar to those for which the network has been trained but which may introduce unforeseen anomalies and variations into the overall task. There is no guarantee that a given set of training data will underwrite an extrapolation to new data sets, but the successful employment of backpropagation techniques for a wide variety of computationally indeterminate tasks (such as visual recognition) has demonstrated the utility of the approach for many contexts of, what at least appears to be, broadly abductive learning. The fact that backpropagation learning algorithms—even of otherwise quite different specifications—are subject to formal *compositionality*, as explicated in the framework of category theory in [30,31], is especially relevant here because it indicates how a variety of such learners may be connected with one another in series and in parallel to collaborate on larger tasks.

The present concern is not of the details of the algorithms themselves, but with certain high-level features that all such backpropagation learning algorithms share. For the purposes of the present argument, one point worth noting initially is that what are called machine *learners* would perhaps be better described as machine *error-correctors*. It is not that these algorithmic architectures are improperly said to learn; they certainly do learn in some sense. However, what is important about their processing is not the fact that they learn, so much as the way in which they learn, namely by systematically adjusting for discrepancies between the output values they generate and the target values specified by a supervising agent. For machine learning algorithms, the supervisor plays the role of authoritative representative for recalcitrant reality that empirical data plays for recollective spirit in the Hegelian/Brandomian account. In both cases, there is a hard limit external to the learning process that authoritatively constrains it. Yet, one (the training supervisor) is itself social and cognitively invested; while the other (physical reality) cannot be said to possess inherent social and cognitive interests. This difference is important because collaborative scenarios in which A.I. machine learning agents and human recollective agents work together will typically involve both types of constraints.

The main claim of the previous section was that Brandom's interpretation of the process of Hegelian recollection may be understood as the characterization of a social and tradition-constituting type of abductive inference. We can now compare and contrast that characterization of recollection with the structure of machine learning backpropagation as just sketched, thereby providing further information. The second core claim of the present argument consists of two parts. Firstly, (A) that there are nontrivial formal parallels between backpropagation algorithms and recollective abduction. Secondly, (B) that despite such parallels, there are structural reasons intrinsic to current methods of machine learning that preclude characterizing machine learning as properly recollective in Brandom/Hegel's sense. Taken together, the two claims (A) and (B) provide good reason to employ the difference between recollective and non-recollective processes as a line of demarcation in human–A.I. collaborative scenarios. Rather than merely invoking the external difference between human and A.I. agents (which presumes as already understood what makes these two types of agent distinct) or, on the other hand, appealing to some more or less mysterious property such as consciousness or moral feeling (which again would presume to explain by way of something that itself stands in need of explanation), the usage of the recollective/non-recollective difference for discriminating agential roles in a given complex, collaborative process of reasoning has the advantage that this difference is subject to clear and objectively determined criteria. Whatever the basis for deciding upon distributions of responsibility and authority in a community, the more definite the distinctions involved in applying the basis, the less subject it will be to arbitrary appropriation and misuse.

With respect then to the first part (A) of the claim, it may seem a bit of a stretch—to say the least—to see in the respective structures of Brandom's abductive recollection and the backpropagation method of machine learning more than a rough analogy. Certainly, it is possible to find a number of structural parallels between the two processes, as both are conceived as epistemically progressive procedures based on a series of "errors"; and both are explicitly sequential in the sense that definite occasions and the resulting discrepancies initiate well-defined operations resulting in definite outputs. Nonetheless, is it not the case that relying on such analogical structures is notoriously problematic, particularly in a field such as artificial intelligence? Indeed, it would be argumentatively suspect to build a positive account on such a basis, for instance to try to use such formal parallels to ground a claim that machine learning algorithms might be particularly apt for this sort of cognitive task. More to the point for this type of question would be research such as that presented and discussed in [32–35], which investigates the prospects and explores the limits of performing abductive inferences computationally. However, the purpose here is not to establish a positive claim; rather, the point is to determine a definite limit for machine learning algorithms of this type, whatever their potential for simulating or enacting such kinds of reasoning. The formal similarities between the two processes serve here only to establish a possible and still indeterminate commonality, with respect to which a kind of fixed upper bound is proposed (B) as the main theoretical result.

What is this upper bound? It is precisely the recollective moment of the process of recollective abduction. However sophisticated the implementation of a neural network architecture with the method of backpropagation might be, the ultimate outcome of the iterative process of training is a single set of weights assigned to the neurons in the network. Even if the results of previous iterations are preserved in memory storage, there is no protocol internal to the backpropagation process itself that would support the self-reflexive narrativization of the process of learning as a process of progressively having come to understand common underlying conceptual content in a way that is better by virtue of having passed through that very process itself.

That the recollective moment of the process is absent from the machine learning method of backpropagation in no way entails that the latter is nonetheless abductive in form and lacking *only* recollection. As suggested above, the overall learning process of backpropagation is in many respects more inductive than abductive. At any rate, two main questions appear to still remain unresolved. First, it is far from clear that the implementation of a backpropagation algorithm may be said to formulate any hypothesis whatsoever with regard to the discrepancy between output and supervised correct value. Thus, the third phase in the schema may not be satisfied. Second, even if the third phase *is* satisfied and some hypothesis H may be identified, it is not evident how the process of machine learning could or should be characterized as discharging H in its own distinctive manner. Thus, the fourth phase in the schema may remain unsatisfied or underdetermined even if the third phase is not in question.

In fact, there does appear to be a reasonable proposal for accounting for the third phase and the characterization of the hypothesis H. This would be to identify H with the set of updated weights generated by the backpropagation algorithm itself at each stage. In other words, each pass of the algorithm would be understood to recalibrate the network as a whole into the form of a newly revised hypothesis. In this way, each iteration of the algorithm (forward pass/backward pass) would result in a definite hypothesis H, namely the set of weights output by that iteration. Each result does, from this (external) point of view, seem to have a broadly abductive character, even if the result itself is entirely determined by the mathematics of the algorithm itself. The epistemic "success" of each iterative pass through the algorithm is contingent upon features of the dataset used by the supervisor as well as features of the network architecture of the learner. Although the descent along the gradient of error is a reliable method for approximating the intended output values, the improvement of the network *even for the fixed supervised values* is by no means guaranteed, much less when extended to new data sets.

If, nevertheless, the set of weights at each iterative stage of the algorithm is taken to be the relevant hypothesis H, the interpretation of the discharge of H would seem to be forced. The discharge of the hypothesis is taken to be the "pragmatic *program* that guides future action in some respect relevantly controlled by H." However, in the context of the implementation of the backpropagation algorithm, such a program can be nothing other than the backpropagation algorithm itself. Simply put, the process as a whole does nothing else with the set of weights assigned at any given stage than operate with them on the basis of the input data provided. In this respect, there is no other kind of epistemic state internal to the machine learning system than "hypotheses". If a hypothesis for a given epistemic agent cannot be meaningfully contrasted with something else, for example knowledge, belief, or disbelief, then it remains dubious whether the term hypothesis has, for that agent, any real content.

We thus appear justified in describing the kind of learning that takes place through supervised training and backpropagation, whether or not such learning may in some sense be abductive in character, as *non-recollective learning*. It is taken for granted that the implementation of machine learning algorithms under suitable conditions of supervision do indeed *learn* in some meaningful sense. Whether the error-correcting process of learning is rightly categorized as abductive, however, remains an open question, although it does seem reasonable to isolate the fourth phase of DISCHARGE as the locus of any proper settling of the matter. In any case, the learning processes organized through algorithmic backpropagation, taken as such and without any additional processing structures through which they might be articulated, remain incapable in principle of performing recollection in the Hegelian sense as analyzed by Brandom in [1]. They simply do not have the requisite referential capacity towards their own learning process that would allow them to articulate not just what they know but how they have come to know it.

4. Abductive Inferences in Mixed Communities

With the difference established between processes of *recollective abduction* on the one hand and of *non-recollective learning* on the other, it becomes possible to pose a number of questions regarding the ways these two types of processes might interact in various contexts of inquiry and potential collaboration. One of the most significant challenges facing artificial intelligence research and development today is not just the technical matter of finding better optimization algorithms or more efficient hardware/software integration, but rather the social and inevitably political question of how human beings and artificial intelligent "agents" of various types can and should be conditioned to interact.

With the results of the previous two sections in mind, it is now possible to formulate a relatively precise question: How should human-AI communities formed via the interactions between (human) processes of recollective abduction and (AI) processes of non-recollective learning be understood and regulated? We will call a community consisting of both agents responsible for recollective abductive inferences and agents who are capable only of non-recollective learning a *mixed community*. It is presumed that such a community organizes itself around practical tasks that require the input and expertise of both types of agent. In other words, neither type of agent is entirely dispensable, given the community's own particular needs.

With regard to the challenges and positive possibilities of mixed communities under such conditions, we present the following results:

1. The self-reflexive dimension of communities of recollection introduces an intrinsic asymmetry between recollective agents and non-recollective learners in mixed communities. Even if non-recollective learners bear an outsized responsibility for generating theoretical results in whatever context, it will ultimately be the responsibility of recollective agents and sub-communities to generate the narratives by means of which those theoretical results are brought into the reflexive self-understanding of the community as a whole.
2. The boundary that determines the inclusion and exclusion of agents from the recollective sub-community by means of mutual recognition is constructed and maintained by that community itself. Thus, any external petition for inclusion in the process (and thereby community) of recollection remains wholly dependent upon the community to which the petition is made. No third-party adjudication is available, even in principle, since there does not appear to be any independent criterion (such as, for example, a sufficiently reliable history of past epistemic success) that would be capable of deciding the matter on grounds not determined by decisions constituting the prior recollective community itself. Nonetheless, the very requirement that the recollective community must explicitly articulate the reasons (and thus the normative criteria) according to which it progressively redefines itself in abductive response to its experiences of error ensures that petitions for inclusion that invoke those very norms definitive of the community that possess a *prima facie* plausibility and thus, by the community's own lights, ought to be at least given fair consideration. Yet, interestingly, the mere consideration of any such petition as worthy of deliberation and response already involves an at least partial recognition of the petitioner as a genuine interlocutor and therefore, already a virtual member of the recollective community itself. The very logic of recollective reason guarantees that any explicit appeal to rational norms and argumentative criteria places the very agent who makes such an appeal at the very least *on the boundary* of the community at stake, if not necessarily already on the hither side of that boundary.
3. The recollective process suggests a strategy for enriching backpropagation algorithms such that they would consist of a two-level architecture that would add (at the first or base level) a second level of expressive elaboration to the backpropagation process that would, at each pass through the algorithm at the lower level, provide an appropriately explanatory account of why and how the descent along the error gradient at that particular stage constitutes an epistemic gain. Techniques currently being developed under the headings of Explainable A.I. (XAI) and Interpretable A.I. (see the overview in [36]) show significant promise in this regard. Yet, it is important to note that at this second level, it would be insufficient for a machine learner merely to report, in however detailed a fashion, the data of what has changed in the internal network. Beyond a mere report, the second level of explanatory elaboration would have to provide an account of *what* was learned and *how* it was learned that, according to some relevant measure, would exceed the information available to a mere inspection of the process itself by translating that information in a constructive way into a genuinely abductive proposal formulated and communicated by the learner itself. While such a machine learner would use a backpropagation algorithm as part of its constitutive learning process, the process itself could not be properly described simply as backpropagation. Instead, the learner would be identified by the enriched algorithmic procedure including the second explanatory and communicative layer. It would not be a backpropagation learner but rather a recollective/backpropagation learner and interlocutor. It remains an open question how exactly the relevant criteria to distinguish insufficiently abductive "mere reporting" from "genuine abductive inference" would be fully determined. However, it is clear that only recollective communities are in a position to delimit and implement such criteria.

5. Conclusions

To take stock of the issues canvassed here, it is relevant that the question of whether non-recollective learning by way of backpropagation algorithms might nonetheless perform abductive inferences remained open in our discussion. The distinction between recollective abduction and non-recollective learning does not draw an entirely precise line because it leaves the status of the category of non-recollective abduction unresolved. In principle, a non-recollective learner or a community of such learners might be capable of performing abductive inferences without simply in virtue of that fact constituting a recollective community, a tradition. This means that it is not foreclosed in advance that mixed communities might collaborate abductively on various tasks without necessarily bringing all the abductive processes involved into the unified framework of recollection. The distinction between the hypotheses conjectured in abductive reasoning and the explanatory narratives that justify such hypotheses as markers of epistemic progress becomes particularly salient. This distinction characterizes the sub-community of those who are capable of deliberating about and collectively constructing narratives of epistemic progress as the appropriate seat of authority and responsibility for the coordinated deployment of the community's reasoning tasks and projects taken as a whole.

The question of how to distribute responsibility and authority in mixed communities is one of the most pressing issues for regulating the roles of computational agents in public and private life. This is a general social and political problem for which the category of *explanation* is and will be essential. For instance, the European General Data Protection Regulations recommend in Recital 71 that when automated procedures result in a decision that affects human individuals, those individuals have a right to be provided with an explanation of how the given decision was generated. See [27] (p. 245). If explanations coming from the automated procedures themselves as non-recollective learners, and potentially abductive agents, are to count as valid for a mixed community, then it is by way of recollective processes that the criteria of such validity must be established. However, this entails that protocols must be established by the recollective sub-communities of the mixed community that specify when and how such explanations are warranted and validated. Again, the recollective/non-recollective distinction is determinative for which agents and sub-communities are invested with authority and responsibility.

We return by way of conclusion, then, to the earlier notion of the i-frame as a necessary condition for processes of recollective abduction. I-frames provide the requisite uncontended structures of deliberation and decision for any community that defines itself self-consciously in terms of a definite communicatively transmitted tradition and *a fortiori* one that constitutes itself by way of processes of recollective abduction. It does not seem that there are any absolute i-frames, although proposals such as that of [37] seem to attempt to find something akin to such absolute structures by isolating transcendental conditions of possibility for rational deliberation as such. The relation between proposals such as [37] and Brandom's Hegelian program of recollective spirit remains a subject for further research. Certainly, any conception of rationality that involves not only providing reasons in support of some given proposal or hypothesis but also explaining why and how the community deliberating over the proposal is better off because of it, will find theoretical resources in Brandom's account of recollection for drawing the necessary distinctions at stake. In any case, in the long run, standards of rationality remain subject to comprehensive revision, and all institutions and conventions will eventually be up for grabs and subject to reconstruction. The human community marks out a vast and vastly complex territory, and it is clear that at least for the near future any mixed community in the sense defined above will form a local and relatively well-circumscribed region within it. Given this situation, and given the result obtained above that places the burden of regimenting relations between recollective and non-recollective agents on the shoulders of the recollective sub-communities, it would seem to be propitious for such recollective agents and their communities faced with this task to take as an initial direction for allocating resources and formulating strategies the planning and tentative articulation of i-frames designed specifically for such occasions.

This argument also shows indirectly the reason as to why, if machine learning processes of whatever type reach a degree of sophistication at which they may properly be said to perform recollective abductions, such a development would herald important consequences for what human–AI communities might then subsequently become. In particular, the development of multi-level machine learning architectures (already a burgeoning research program for various other reasons) that explicitly narrate at level $n + 1$ (by text, flowchart or in some other manner) what at level n is learned in subsequent stages, presents itself as a foray into the space of options and challenges made available by the possibility of recollective A.I.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Brandom, R. *A Spirit of Trust*; Belknap/Harvard University Press: Cambridge, MA, USA, 2019.
2. Hegel, G.W.F. *The Phenomenology of Spirit*; Miller, A.V., Translator; Oxford University Press: Oxford, UK, 1977.
3. Brandom, R. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*; Harvard University Press: Cambridge, MA, USA, 1994.
4. Brandom, R. *Tales of the Mighty Dead: Historical Essays in the Metaphysics of Intentionality*; Harvard University Press: Cambridge, MA, USA, 2002.
5. Carnap, R. *The Logical Syntax of Language*; Smeaton, A., Translator; Open Court: Chicago, IL, USA, 2002.
6. Wittgenstein, L. *Philosophical Investigations*; Anscombe, G.E.M., Translator; Blackwell: Oxford, UK, 1978.
7. Sellars, W. *Pure Pragmatics and Possible Worlds: The Early Essays of Wilfrid Sellars*; Sicha, J., Ed.; Ridgeview: Atascadero, CA, USA, 2005.
8. McDowell, J. *Mind and World*; Harvard University Press: Cambridge, MA, USA, 1994.
9. Pippin, R.B. *Hegel on Self-Consciousness: Desire and Death in the Phenomenology of Spirit*; Princeton University Press: Princeton, NJ, USA, 2011.
10. Peirce, C.S. *The Essential Peirce: Selected Philosophical Writings*; 2 Volumes; Peirce Edition Project, Ed.; University of Indiana Press: Bloomington, IN, USA, 1998.
11. Aliseda, A. *Abductive Reasoning: Logical Investigations into Discovery and Explanation*; Springer: Dordrecht, The Netherlands, 2006.
12. Gabbay, D.M.; Woods, J. *The Reach of Abduction: Insight and Trial*; Elsevier: Amsterdam, The Netherlands, 2005.
13. Magnani, L. *Abductive Cognition: The Epistemological and Eco-Cognitive Dimensions of Hypothetical Reasoning*; Springer Press: Berlin, Germany, 2009.
14. Park, W. *Abduction in Context: The Conjectural Dynamics of Scientific Reasoning*; Springer: Berlin, Germany, 2017.
15. Restrepo, J.A.F. *Are There Types of Abduction? An Inquiry into a Comprehensive Classification of Types of Abduction In Abduction in Cognition and Action: Logical Reasoning, Scientific Inquiry and Social Practice*; Shook, J., Paavola, S., Eds.; Springer: Berlin, Germany, 2021; pp. 3–30.
16. Shook, J. *Abduction in Cognition and Action: Logical Reasoning, Scientific Inquiry and Social Practice*; Paavola, S., Ed.; Springer: Berlin, Germany, 2021.
17. Sellars, W. Some Reflections on Language Games in Science. In *Perception and Reality*; Ridgeview Publishing: Atascadero, CA, USA, 1991; pp. 323–358.
18. Brandom, R. *Between Saying and Doing: Towards an Analytic Pragmatism*; Oxford University Press: Oxford, UK, 2008.
19. Gangle, R.; Caterina, G.; Tohmé, F. Abductive Spaces: Modeling Concept Framework Revision with Category Theory. In *Abduction in Cognition and Action: Logical Reasoning, Scientific Inquiry and Social Practice*; Shook, J., Paavola, S., Eds.; Springer: Berlin, Germany, 2021; pp. 49–73.
20. Brandom, R. A Hegelian Model of Legal Concept Determination: The Normative Fine Structure of the Judge's Chain Novel. In *Pragmatism, Law, and Language*; Hubbs, G., Lind, D., Eds.; Routledge: New York, NY, USA, 2014; pp. 19–39.
21. Hegel, G.W.F. *Philosophy of Right*; Dyde, S.W., Translator; Prometheus: Amherst, NY, USA, 1996.
22. Luhmann, N. *Organization and Decision*; Barrett, R., Translator; Cambridge University Press: Cambridge, UK, 2018.
23. Winfield, R. Hegel, Mind, and Mechanism: Why Machines Have No Psyche, Consciousness or Intelligence. *Hegel Bull.* **2009**, *30*, 1–18. [CrossRef]
24. Van Tuinen, S. Philosophy in the Light of A.I. *Angelaki* **2020**, *25*, 97–109. [CrossRef]
25. Hui, Y. *Recursivity and Contingency*; Rowman and Littlefield: London, UK, 2019.
26. Negarestani, R. *Intelligence and Spirit*; Urbanomic: Falmouth, UK, 2018.
27. Kelleher, J.D. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2019.
28. Wilmott, P. *Machine Learning: An Applied Mathematics Introduction*, 2nd ed.; Panda Ohana Publishing: Coppell, TX, USA, 2020.

29. Deisenroth, M.P.; Faisal, A.A.; Ong, C.S. *Mathematics for Machine Learning*; Cambridge University Press: Cambridge, UK, 2020.
30. Fong, B.; Spivak, D.; Tuyeras, R. Backprop as Functor: A compositional perspective on supervised learning. *arXiv* **2017**, arXiv:1711.10455.
31. Spivak, D. Learners' Languages. *arXiv* **2021**, arXiv:2103.01189.
32. Mooney, R.J. Integrating Abduction and Induction in Machine Learning. In *Abduction and Induction in Artificial Intelligence*; Flach, P., Kakas, A., Eds.; Kluwer: Dordrecht, The Netherlands, 2000; pp. 181–191.
33. Denecker, M.; Kakas, A. Abduction in Logic Programming. In *Computational Logic (Kowalski Festschrift)*; Kakas, A., Sadri, F., Eds.; Springer: Berlin, Germany, 2002; pp. 402–436.
34. Bylander, T.; Allemang, D.; Tanner, M.C.; Josephson, J.R. The computational complexity of abduction. *Artif. Intell.* **1991**, *49*, 25–60. [CrossRef]
35. Ignatiev, A.; Narodytska, N.; Marques-Silva, J. Abduction-based explanations for Machine Learning models. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 1511–1519.
36. Linardatos, P.; Papastefanopoulus, V.; Kotsiatis, S. Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy* **2020**, *23*, 18. [CrossRef] [PubMed]
37. Habermas, J. *The Theory of Communicative Action*; 2 Volumes; McCarthy, T., Translator; Beacon Press: Boston, MA, USA, 1987.