THE ROLE OF REFLEXIVITY IN UNDERSTANDING HUMAN UNDERSTANDING

Steven James Bartlett

Website: http://www.willamette.edu/sbartlet

· · · · ·

THE ROLE OF REFLEXIVITY IN
UNDERSTANDING HUMAN UNDERSTANDING

STEVEN J. BARTLETT

Philosophy is reflective. The philosophizing mind never simply thinks about an object, it always, while thinking about any object, thinks also about its own thought about the object. Philosophy may thus be called thought of the second degree.

Philosophy . . . has this peculiarity that reflection upon it is part of itself. . . . [T]he theory of philosophy is itself a problem for philosophy; and not only a possible problem, but an inevitable one.

— R. G. Collingwood [1]

THE INTERNAL LIMITATIONS OF
HUMAN UNDERSTANDING

We carry, unavoidably, the limits of our understanding with us. We are perpetually confined within the horizons of our conceptual structure. When this structure grows or expands, the breadth of our comprehensions enlarges, but we are forever barred from the wished-for glimpse beyond its boundaries, no matter how hard we try, no matter how much credence we invest in the substance of our learning and mist of speculation.

The limitations in view here are not due to the mere finitude of our understanding of ourselves and of the world in which we live. They are limitations that come automatically and necessarily with *any* form of understanding. They are,

---

[1]   *The Idea of History* (London: Oxford University Press 1946), p. 1, and *An Essay on Philosophical Method* (London: Oxford University Press 1933), pp. 1-2.

as we shall see, part and parcel of any organization or ordering of data that we call information.

The consequences of these limitations are varied: As a result of them, hermeneutics cannot help but be hermetic; scientific theories of necessity are circumscribed by the boundaries of the ideas that define them; formal systems must choose between consistency and comprehensiveness; philosophical study, because it includes itself within its own proper subject matter, is forced to be reflexive in its self-enclosure. The fundamental *dynamic* shared by all forms of understanding testifies to an internal limitative keystone.

Kant's architectonic suggested the existence of this keystone, which was expressed in his theory of subjective constitution, the molding of the world by the built-in categories of human intelligibility. Whorf's study of natural languages sought to make this keystone apparent in his theory of linguistic relativity, which proposed that what we can grasp is limited by the expressive capacities of our language. Gödel, Löwenheim, Skolem, and others felt its presence in the world of formal proof, in a variety of forms that recapitulate linguistic relativity on a formal level. Husserl and his student, Eugen Fink, seemed to recognize its reality in the self-contained nature of the phenomenological attitude, which requires a basic leap from an intuitive, "naturalistic" understanding of the world, to a conceptual *conversion* that brings with it an essentially distinct approach to self-understanding. In large- and in small-scale physics, aspects of the same limitative dynamic are visible in both relativity theory and quantum mechanics, in the form of systematic acknowledgements that physical reality is intrinsically defined as a function of the observer's framework and state.[2]

Not by any concerted act of imagination can we trespass beyond the boundaries of what for us is imaginable. — This is a tight tautology, within which we realize all the freedom that is possible for us. The internal limitations of human understanding disclose themselves in several distinct ways: In our practical dealings with the world, we are subject to *neurological* limitations and to limitations of *language* and *idea*. And in our conceptual efforts, we are constrained by *epistemological* boundaries.

The limitations that structure our practice are set in place, and yet also are revealed to us, by human neurology, by the range of concepts available to us, and in part by the structure of human natural and formal language. Our neurol-

---

2   These and other examples are discussed in the author's introductory essay, "Varieties of Self-Reference," in Steven J. Bartlett and Peter Suber, eds., *Self-Reference: Reflections on Reflexivity* (Dordrecht, Holland: Martinus Nijhoff 1987).

ogy, conceptual vocabulary, and linguistic resources all are *encoding systems* that provide us with the spectrum or palette of colors in terms of which the world we inhabit develops its reality. Human neurology defines what we are able to apprehend and to which we can respond; the range of our concepts and the structure of our language enable us to think and to talk — within the elastic boundaries we must ever carry with us. Beyond these limitations ingredient in our encoding *abilities*, there are certain epistemological boundaries, which we will touch upon in a moment, that define the limits of *possible knowledge*.

The picture of the human condition suggested by these limitative factors is one of a finite organism whose neurology is responsive to a range of possible stimuli, whose conceptual vocabulary permits a certain breadth of theoretical representation, whose natural and abstract languages allow for a scope of expression and demonstration, and whose extent of knowledge is determined by conditions and limits described by epistemology. This is a picture of a creature who inhabits a specifically human universe of meaning, one that seems to be a fragment — a larger of a smaller fragment, but a fragment nonetheless — of a more inclusive reality, from contact with which our practical and theoretical limitations eternally bar us: what has, in short, been called "noumenal reality."

Appealing though this picture may be to poetic inspiration, it is a grossly distorted one: It misconstrues the compass and the kind of internal limitation that is our subject here. This view, which situates human reality within a more comprehensive framework, *exports* and yet *presupposes* the very concepts, language, and neurology that define the human perspective. In this step of exportation, we run headlong into the invisible netting of epistemology's constraints, from which "escape" is not only impossible but, on reflection, also is unthinkable. The existence of these constraints is theoretically determined, and does not depend upon the contingent biological, conceptual, or linguistic abilities of a particular organism: In attempting to refer beyond the reality made possible by our neurology, concepts, and language, we attempt, in essence, to refer beyond the reach of our referring capacities. We seek to do the impossible — not the impossible in practice — but the impossible in principle.

The so-called "boundaries" of our understanding are very peculiar limits, unlike the boundaries that delimit a field, or the walls that enclose a box. They much more closely resemble the self-limiting and yet unbounded character of a continuum that has no "outside," such as is formed by a topologically recurved surface or volume. A close analogy is the relativistic model of the physical universe, unbounded yet finite. In such a model, no matter where one goes, no matter how far, there is no way "out." For the very notion of an "outside" is *part* of the universe of meaning whose internal limitations we may now perhaps begin to appreciate. These "limitations" are of a special, philosophical variety; here the ordinary meaning of the word has undergone a radical change.

If we cannot reasonably assert (or deny) that there is an "outside," lying "beyond the reach" of the powers of our neurologies, concepts, and languages, then does it in fact make sense to say that we are *constrained* by these alleged internal limitations? Where do these limits, which we cannot meet, touch, or see, reside? Is it merely a Procrustean stretching of language to suggest that these are "limits" at all?

SELF-REFERENCE AS A TOOL OF CONCEPTUAL ANALYSIS

Internal limitations of this kind cannot by direct assault be coerced to show themselves. Yet their presence will be made evident to us, again and again, and often with little effort, if we employ an indirect strategy.

Hints of the internal limitations of human understanding came first in the guise of paradoxes involving self-reference. For centuries, these were dismissed as sophistry, no more than interesting philosophical parlour tricks. But then they became serious when suddenly, at the turn of the century, a number of paradoxes in number theory and the theory of classes were produced through the use of reflexive strategies. Other paradoxes were soon discovered, usually through the reflexive application of certain inconsistency-engendering predicates. Within a dozen years this family of paradoxes came to include the Burali-Forti paradox (1897), Cantor's paradox (1899), Russell's paradox (1901), the Richard paradox (1905), the Zermelo-König paradox (1905), Berry's paradox (1908), Grelling's paradox (1908), followed by others.

For a time, the fact that reflexive or self-referential techniques could lead to paradox brought criticism to bear on the use of self-reference itself: If we simply shunned reflexivity, we might be spared the intellectual inconvenience of paradox.[3] But the phenomenon refused to recede off stage.

Beginning in 1931, when Kurt Gödel's famous paper on formally undecidable propositions was published, a rash of results broke out, all relating to the discovery of internal limitations of formal deductive systems. Again, the main tools used were reflexive. Numerous theorems of formal limitation were proved — by Kleene, Roser, Kalmár, Gentzen, Church, Turing, Post, Tarski, Mostow-

---

3   For the present, the terms 'self-reference' and 'reflexivity' are used interchangeably. As will be seen in this volume, contributors to the literature differ in their preference for one term or the other; often 'reflexivity' appears to be the more general term, but no consensus has formed.

ski, Löwenheim, Skolem, Henkin, Wang, Curry, Myhill, Chwistek, Uspenskij, Kreisel, and others.

The ripples from these reflexively acquired results quickly spread to the then-new field of cybernetics, and, in time, on to its most recent progeny, among them, systems theory, information theory, and artificial intelligence. The central ideas of positive and negative feedback and feedforward were developed and applied to a growing range of topics, from research concerned with self-regulating and self-correcting systems, to studies of the human brain, psychotherapeutic interventions, and biological homeostasis.[4]

In retrospect, the self-referential techniques used in mathematical logic and the foundations of mathematics were employed in what seems to have been an almost intuitive fashion. The reflexive strategies they exemplified have had to wait for a metatheory to clarify the underlying unselfconscious practice. To some extent, reflexive studies and applications in philosophy, which began to flourish in the two, three, and four decades following the discovery of the formal paradoxes, were more methodologically self-aware.

In philosophy, the phenomenon of self-reference has inspired research in three main areas: in semantic theory, theory of argument, and theory of knowledge. Of these, the earliest studies of reflexivity were made in semantic theory. They sought to understand the impact of the paradoxes encountered in number theory and the theory of classes upon the capacity of propositions, both those of formal systems and those in non-formalized discourse, to assert truth without self-referential inconsistency. Papers in Part I of this collection share this focus.

Somewhat later, a small group of philosophers began to cultivate an explicit interest in the use of self-reference in philosophical argument. Although individual examples of reflexive argumentation have peppered the history of philosophy, it was not until the middle of the twentieth century that efforts were made to construct a theory of self-referential argumentation, by examining a specific variety of self-reference that has come to be called "pragmatical" or "performative." Papers in Part II study this topic.

A third area of philosophical interest inspired by reflexivity evolved from the Kantian and Husserlian attempts to identify the transcendental preconditions of objective knowledge. Here, the internal limitations of human self-understanding become especially evident, in the human effort to acquire knowledge

---

4   An extensive bibliography of more than 1,200 works relating to self-reference, prepared by Peter Suber, will be found in Bartlett and Suber, *op. cit.*

about the limits of what human subjects can know. The variety of self-reference relevant to this task has come to be called "metalogical." Studies of metalogical self-reference describe the general and necessary conditions that underlie our abilities, in principle, to refer at all, no matter what the object may be to which reference is made. Papers in Part III relate to with this area of study.

A fourth area of interest in reflexivity, closely allied to the three just mentioned, has recently developed in artificial intelligence. Here, philosophy has continued a long tradition as mother to a succession of disciplines: From a historical point of view, investigations of computational reflexivity in artificial intelligence grew out of studies of self-reference in formal systems — undertaken by researchers in mathematical logic, foundations of mathematics, and semantic theory, disciplines all of philosophical origin.

Research in artificial intelligence attempts to simulate certain human abilities (at present, more easily formulated, elementary abilities), which frequently are reflexive in nature, in a context in which computational capacities can surpass in speed and complexity those of their human creators. Studies of reflexivity in artificial intelligence have sought, for example, to develop computer programs enabling a non-human electronic system to coordinate facts, establish connections among them, and on this basis to generate logically necessary, or plausible, inferences about the world. There has been a growing realization among researchers that such computational languages indeed must themselves constitute reflexive representations of reality, since the representations they make possible form part of the reality to be understood; and so we again encounter our topic, in another form. A programming language capable of general, reflexive intelligence immediately poses the need for self-referential abilities, to allow a machine to reflect on the usage of the language by the very machine whose functioning is defined by it. Papers in Part IV are variously concerned with problems in this area, and describe several ground-breaking proposals.

Finally, papers in Part V illustrate applications of a number of techniques of self-referential argumentation.

Studies of the various forms of reflexivity — semantical, pragmatical, metalogical, and computational — have contributed, as papers in this volume will make clear, to the task of making the intangible limits of understanding more clearly manifest to us who are constrained by them.

THE CRITICAL AND
CONSTRUCTIVE USES OF SELF-REFERENCE

In keeping with philosophy's first commission, to provide a critical propaedeutic to self-assured clarity, philosophical applications of self-reference, on which

I will now focus, have tended for the most part to be critical, negative, or corrective, seeking to identify and eliminate internally inconsistent dogmas. Some applications, however, have sought to use self-referential approaches to establish results non-destructively.

PRAGMATICAL SELF-REFERENCE

[M]y argument does not, at least in any obvious way, miss the point of anyone who might contend that philosophical statements can be true or false independently of the arguments used to establish or disestablish them. It acquires its force precisely from the force of this contention; for the contention can only take the form of an argument, and this very argument will at once serve as a further illustration of the thesis I have been advocating.

— Henry W. Johnstone, Jr.[5]

Pragmatical self-reference directs attention to the factual commitments involved in making an assertion. For example, the assertion, "knowledge is impossible in this world of flux," is pragmatically self-referentially inconsistent: Provided that the assertion is in fact linked to an underlying commitment that places it in the category of knowledge-claims, the assertion is self-falsifying. The challenging task of the pragmatical self-referential analyst is to reveal the existence of the factual commitments that underlie everyday and philosophical discourse. His results stand or fall depending on the convincingness of his factually-focused demonstration.

THE CRITICAL USE OF
PRAGMATICAL SELF-REFERENCE

Every philosophical system is subject to the obligation of accounting for its own possibility; it must at least be able to give such an account in its own terms. Less radically expressed, there must be no incompatibility between the doctrinal content of a philosophical theory, that which is maintained and asserted in it, on the one hand, and, on the other, the mere fact of the formulation of the theory in question. An incompatibility of such a kind

---

5   Henry W. Johnstone, Jr., *Philosophy and Argument* (University Park, Pennsylvania: Pennsylvania State University Press 1959), p. 81.

would provide the basis for a decisive argument against the theory beset by that incompatibility.

— Aron Gurwitsch[6]

The strange thing is that philosophers should have been able to hold sincerely, as part of their philosophical creed, propositions inconsistent with what they themselves knew to be true; and yet, as far as I can make out, this has really happened.

— G. E. Moore[7]

Pragmatical applications of self-reference have attempted to show that such claims as these are self-falsifying:

- Pleasure is the chief good, since any good thing is made more desirable by the addition of pleasure.[8]

- The materialist can explain the causes of our ideas in terms of external bodies.[9]

- Every event must have a cause.[10]

- All knowledge, including this, is a product of an organism's adjustment to its environment.[11]

- All meaningful statements are verifiable.[12]

---

6   Aron Gurwitch, "An Apparent Paradox in Leibnizianism," *Social Research*, Vol. 33, No. 1, 1966, p. 47.

7   G. E. Moore, "A Defense of Common Sense," *Classics of Analytic Philosophy*, ed. R. R. Ammerman (New York: McGraw-Hill 1965), pp. 53-54.

8   Argument from Eudoxus; see treatment by H. W. Johnstone, Jr., *op. cit.*, pp. 64ff.

9   H. W. Johnstone, Jr., *op. cit.*, pp. 67ff.

10   Argument from Hume; see discussion in Johnstone, *op. cit.*, p. 95.

11   W. M. Urban, *Beyond Realism and Idealism* (London: Allen and Unwin 1949), p. 236, and the discussion in Johnstone, *op. cit.*, pp. 69ff.

12   Richard Rorty, "The Limits of Reductionism," in J. Lieb, ed., *Experience, Existence, and the Good* (Carbondale: Southern Illinois University Press 1961), pp. 100-116; cf. esp. pp. 104-107.

- Science is incapable of objectivity.[13]
- The shift from one theory to another involves an incommensurable change in the meanings of the terms used, so that there cannot be any statements invariant across theories.[14]

- No hypothesis can be immune to revision.[15]

- No hypothesis can be irrevocably falsified.[16]

- All our statements lack significance.[17]

To this short list could be added many other examples, for numerous philosophical positions have been indicted for falling victims to the pragmatical variety of self-referential inconsistency. Among those that have been attacked in this way are the coherence theory of truth,[18] pragmatism,[19] scepticism,[20] intuitionism,[21] behaviorism,[22] determinism,[23] subjectivism,[24] views that oppose idealism,[25] and views that oppose utilitarianism.[26]

---

13   Carl R. Kordig, "Objectivity, Scientific Change, and Self-Reference," in *Boston Studies in the Philosophy of Science*, Vol. 8 (Dordrecht, Holland: R. Reidel 1970), pp. 519-523.

14   *Ibid.*

15   *Ibid.*

16   *Ibid.*

17   John Passmore, *Philosophical Reasoning* (London: Duckworth 1961), p. 69.

18   E. G. Spaulding, *The New Rationalism* (New York: Holt 1918), pp. 350-351.

19   Josiah Royce, "The Eternal and the Practical," *Philosophical Review*, Vol. 13, 1904, pp. 128-129.

20   W. M. Urban, *The Intelligible World* (London: Allen and Unwin 1929), pp. 45-46, and John Passmore, *op. cit.*, pp. 72ff.

21   W. E. Hocking, *Types of Philosophy* (New York: Charles Scribner's Sons 1939), p. 201.

22   A. O. Lovejoy, "The Paradox of the Thinking Behaviorist," *Philosophical Review*, Vol. 31, 1922, pp. 142-147.

23   J. R. Lucas, *Freedom of the Will* (Oxford: Clarendon Press 1970) and J. M. Boyle, Jr., G. Grisez, and O. Tollefsen, *Free Choice: A Self-Referential Argument* (Notre Dame, Indiana: University of Notre Dame Press 1976).

24   An argument originally advanced by Protagoras: see treatment in John Passmore, *op. cit.*, pp. 64ff.

25   Josiah Royce, *Lectures on Modern Idealism* (New Haven, Connecticut: Yale University Press 1919), pp. 237-240.

26   Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (Oxford: Clarendon Press 1876), Chap. 1, sections 13-14.

THE CONSTRUCTIVE USE OF
PRAGMATICAL SELF-REFERENCE

[V]alid constructive arguments in philosophy must in fact be circular. . . . All valid constructive philosophical arguments involve this element of feedback.

— Henry W. Johnstone, Jr.[27]

In contrast to the preceding critical arguments that utilize pragmatical self-reference to undercut a disagreeable thesis, a few philosophers have tried to use the approach constructively.

(We should remark that this distinction, between critical and constructive arguments, admittedly is often difficult to draw clearly, especially in the present context: A pragmatically critical argument establishing that *P* is self-falsifying leads to the conclusion *not-P*; yet, if *not-P* is thought to be a philosophically significant result, the argument's proponent naturally believes his argument is constructive. Among arguments and their proponents, the constructiveness of their conclusions can be stretched across a broad spectrum. At the dim end of lesser interest one might place, for instance, the critical argument against the assertion, "All our statements lack significance." The self-referential argument that establishes the negation of this assertion resists being thought of as especially interesting or constructive. Certainly it tells us something of which few are ignorant.)

In general, constructive self-referential argumentation attempts to demonstrate a positive thesis, rather than to undermine an erroneous view maintained by someone else: Admittedly, when it comes to showing that others are wrong, the judo-like strategy of utilizing feedback in argumentation is especially well-suited, as a reader new to the field intuitively may suspect. But some constructive arguments have, nevertheless, been formulated using the tools of pragmatical self-reference. A few we might mention here are:

- Moore's defense of common sense, using its appeal;[28]

---

[27]   H. W. Johnstone, Jr., *op. cit.*, pp. 76, 68.

[28]   J. Passmore, *op. cit.*, pp. 78ff.

- The argument that there are invariant conditions of discourse;[29]

- Arguments seeking to demonstrate the ontological commitments of discourse, and the related argument claiming that all objects of which we are conscious are, in diverse senses, real;[30]

- The self-confirming evidence that a sound is audible, *is* that we hear it;[31]

- The defense of "orientational pluralism" in philosophy: According to this view, philosophical positions represent relativistic frames of reference. For them, there is no *unique* solution to philosophical problems.[32]

To these examples may be added the larger group of arguments that progress from a self-referential refutation of an opposing thesis to the affirmation of its philosophically significant negation. Among these are found the positions mentioned earlier that defend: the objectivity of science, free choice, utilitarianism, idealism, the thesis that verifiability is not a property belonging to all meaningful statements, etc.

METALOGICAL SELF-REFERENCE

[W]e are brought to the conclusion that we can never transcend the limits of possible experience.

— Immanuel Kant[33]

Unlike strategies of argumentation using pragmatical self-reference, metalogical approaches direct attention to the commitments that are necessarily in-

---

[29]   J. Passmore, *op. cit.*, pp. 69ff; Paul Lorenzen, *Normative Logic and Ethics* (Mannheim/Zürich: Bibliographisches Institut 1969), p. 14, and (in connection with operative logic) p. 89. See also Lorenzen's *Einführung in die operative Logik und Mathematik* (Berlin: Springer Verlag 1969).

[30]   W. V. O. Quine, *Ontological Relativity* (New York: Columbia University Press 1969), and Alexius Meinong, "The Theory of Objects," in Roderick M. Chisholm, ed., *Realism and the Background of Phenomenology* (Glencoe, Illinois: The Free Press 1960), pp. 76-117.

[31]   Mentioned by J. S. Mill, "Proof of the Principle of Utility," in *Utilitarianism* (Indianapolis, Indiana: Bobbs-Merrill 1971; originally published 1863), Chapter IV; also see discussion in H. W. Johnstone, Jr., *op. cit.*, pp. 77ff.

[32]   Nicholas Rescher, "Philosophical Disagreement: An Essay towards Orientational Pluralism in Metaphilosophy," *Review of Metaphysics*, Vol. 32, No. 2, 1979, pp. 217-251.

[33]   *Critique of Pure Reason*, Preface to the Second Edition, B xix.

volved if a claim or concept *in principle* is to be capable of referring to those objects to which reference is presupposed. Philosophical argument that relies upon principles of pragmatical reflexivity is usually and appropriately considered to be *ad hominem* in nature. Reflexive argumentation developed on a metalogical basis, on the other hand, has an unmistakable transcendental orientation.

Universally, for a claim to function as such it must refer to certain objects, about which assertion is made. Metalogical reflexivity comes to be of interest in connection either, from a critical point of view, with claims that conflict with their own referential preconditions, or, from a constructive point of view, with claims that compel assent, since they cannot be denied without producing such a conflict.

ITS CRITICAL USE

[O]ne must avoid the error of assuming that the sense behind familiar notions is obvious.

— D. C. Ipsen[34]

It constitutes a great advance in our critical attitude. . . to realize that a great many of the questions that we uncritically ask are without meaning. . . . [O]ne is making a significant statement about his subject in stating that a certain question is meaningless.

— P. W. Bridgman[35]

Metalogical applications of self-reference have attempted to identify a wide range of self-undermining concepts and claims. Among these are:

- Descartes' methodologically sceptical hypothesis of an evil genius, capable of shaking all confidence in our abilities to ascertain the truth about reality;[36]

---

[34] D. C. Ipsen, *Units, Dimensions, and Dimensionless Numbers* (New York: McGraw-Hill 1960), p. v.

[35] P. W. Bridgman, *The Logic of Modern Physics* (New York: Macmillan 1961; first printed 1927), pp. 28-29.

[36] O. K. Bouwsma, "Descartes' Evil Genius," in Alexander Sesonske and Noel Fleming, eds., *Meta-meditations: Studies in Descartes* (Belmont, California: Wadsworth Publishing Co. 1965), pp. 26-36; and Steven J. Bartlett, "Hoisted by Their Own Petards: Philosophical Positions that Self-Destruct," *Argumentation*, Vol. 2, 1988, pp. 221-232.

- Kant's distinction between objects spatially structured by the human mind and "objects themselves," to which the human concept of space does not apply;[37]

- The hidden variable interpretation of quantum mechanics, which expressed a bias in favor of realism and physical determinism on the level of small-particle interactions;[38]

- Philosophical scepticism as treated by P. F. Strawson;[39]

- The argument (which ironically depended on a pragmatically reflexive strategy) attempting to show that the *rejection* of free choice is self-falsifying, or else pointless;[40]

- The view claiming that solutions to mathematical or other problems are "discovered"; they are not "invented";

- The opposing view, claiming that solutions to mathematical or other problems are "invented"; they are not "discovered";[41]

- The doctrine that there exists (or does not exist) a "metaphysical self";[42]

---

[37] See Bartlett, *ibid.*

[38] Cf. Steven J. Bartlett, "Self-Reference, Phenomenology, and Philosophy of Science," *Methodology and Science*, Vol. 13, No. 3, 1980, section VII.

[39] See Strawson's argument against scepticism in *Individuals*, which can be interpreted as an attempt to prove that the sceptic's position is metalogically self-undermining.

[40] This argument was advanced in J. M. Boyle, Jr., G. Grisez, and O. Tollefsen in *Free Choice: A Self-Referential Argument* (Notre Dame, Indiana: University of Notre Dame Press 1976). Although a hard-working attempt to show that freedom of choice may be rejected only on pain of pragmatical self-referential inconsistency or pointlessless, the argument itself is metalogically self-undermining. See Bartlett, review of Boyle-Grisez-Tollefsen's book, in *Review of Metaphysics*, Vol. 32, No. 4, 1979, pp. 738-740.

[41] On this hypothesis and the preceding one, see Bartlett, "A Metatheoretical Basis for Interpretations of Problem-Solving Behavior," *Methodology and Science*, Vol. 11, No. 2, 1978, pp. 59-85: esp. pp. 70-72, 79-82.

[42] Steven J. Bartlett, "The Use of Protocol Analysis in Philosophy," *Metaphilosophy*, Vol. 9, Nos. 3 and 4, 1978, pp. 324-336.

- The belief that a phenomenological description of an experience tells us what was "already present" in the experience pre-reflectively and implicitly;[43]

- The Newtonian concepts of absolute time and space;[44]

- The realist view that accords a separable existence to past or future events, independently of the present;[45]

- The framework-independent concept of absolute truth;[46]

- The doctrine claiming that every event is the effect of a prior cause, and the related doctrine claiming that in a cause-effect sequence, the occurrence of the cause was indispensable to the occurrence of the effect;[47]

- The interrelated beliefs that there is a common "pole," called "the ego," shared by all of the investigator's experiences; that consciousness is a universal attribute of experience; that consciousness is a kind of "container" of experiences, beyond which meaningful claims may be made;[48]

- the doctrine that mental events are in many instances the results of prior acts (a belief inspired by the causal dogma mentioned earlier);[49]

- The belief that reflection does (or does not) perturb the structure or nature of pre-reflective experience;[50]

43   Steven J. Bartlett, "Phenomenology of the Implicit," *Dialectica*, Vol. 29, Nos. 2-3, 1975, section III, and "Fenomenologia Tego, Co Implikowane," *Roczniki Filozoficzne*, Vol. XXII, No. 1, 1974, pp. 73-89.

44   Steven J. Bartlett, *A Relativistic Theory of Phenomenological Constitution: A Self-Referential, Transcendental Approach to Conceptual Pathology* (Université de Paris, 1970; *Diss. Abs. Internatl.* No. 7905583), Chapter 2.1.

45   *Ibid.*

46   *Ibid.*, Chapter 2.4.

47   *Ibid.*, Chapter 2.5.

48   *Ibid.*, Chapter 2.6.

49   *Ibid.*, Chapter 2.7.

50   *Ibid.*, Chapter 2.7.

THE CONSTRUCTIVE USE OF
METALOGICAL SELF-REFERENCE

[E]very true proposition attributing a predicate to a subject is purely analytic, since the subject is its own nature.

— Bertrand Russell[51]

The constructive use of metalogical self-reference depends upon a special property of claims of a certain kind: This is the property possessed by a claim that is such that its denial leads exactly to the variety of self-referential inconsistency that is in view here, i.e., self-referential inconsistency that *precludes* that the intended reference of the claim is possible at all.

Claims of this kind are *self-validating*, since, if they are rejected, they succumb to self-referential inconsistency of such magnitude that their capacity to be meaningful is short-circuited. As in the case of pragmatically reflexive arguments, there is an interplay between the critical and the constructive ends to which metalogically reflexive arguments may be put. The relation between criticism and construction is similarly bridged here by a conditional: If it can be shown that a claim is metalogically self-undermining, then the *rejection* of that claim will compel assent. It is important to notice that the rejection of a claim does not entail the positive endorsement of its negation. For example, the rejection of "there exists a metaphysical self" does not commit us to "one does not exist." — Both claims employ a framework-transcending concept that stands in conflict with its framework-relative basis.

In a similar way, the rejection of a self-validating claim is self-undermining.

Among positions and arguments that have sought their own validations in ways closely akin to a metalogically reflexive strategy, these could be listed:

- Kant's transcendental deduction;

- Collingwood's absolute presuppositions of systematic thought, which are presupposed by any cognition, and make knowledge possible;

---

[51] Bertrand Russell, "The Monistic Theory of Truth," in *Philosophical Essays* (New York and London: Longmans, Green, and Co. 1910), p. 167.

- · Husserl's conception of transcendental phenomenology, the analysis of which reflexively discloses the necessary foundation for its own possibility;

- · Strawson's attempt to deduce, in a quasi-transcendental manner, the necessary and basic structure of a conceptual system that makes objective knowledge possible;

- · Gaston Isaye's transcendental method of retortion, which seeks to identify the conditions of the possibility of thought by means of a strategy to show that every possible denial of a self-justifying assertion leads to a self-referentially inconsistent position;[52]

- · The following pair of mutually reinforcing positions: The author's reflexive argument that metalogical referential consistency is a necessary condition of meaning, on the one hand, and his relativistic theory of the constitution of experience, on the other. Together, these approaches show that a wide range of everyday and technical concepts is metalogically self-undermining, underscoring the need for a vocabulary of radically different but referentially self-consistent concepts.[53]

. . .

The experiential and conceptual space we inhabit has a strange structure and a surprising logic that we are only beginning to accept and appreciate. The history of physics serves almost as a parable, for it tells a larger story through having come full circle, from primitive anthropocentrism, to the displacement of man as the center around which all things revolve, to an observer-based awareness of relativity. In this recent return to framework-relativity, self-reference has played an increasingly important role across many disciplines, as a tool of discovery

---

[52]   See Gaston Isaye, "La Justification critique par rétorsion," *Revue Philosophique de Louvain*, Vol. 52, 1954, pp. 205-233; Otto Muck, *The Transcendental Method*, trans. by William D. Seidensticker (New York: Herder and Herder 1968), pp. 163-180; and Martin X. Moleski, "Retortion: The Method of Gaston Isaye," *International Philosophical Quarterly*, Vol. 17, No. 1, pp. 59-83. (Moleski refers Isaye's reflexive approach to the category of pragmatical self-referenc , what is clearly a mistake, given Isaye's self-conscious focus on necessary, non-contingent truths that are conditions of the possibility of human thought.)

[53]   Cf., *inter alia*, Steven J. Bartlett, "Referential Consistency as a Criterion of Meaning," *Synthese*, Vol. 52, 1982, pp. 267-282, reprinted in this volume; and *A Relativistic Theory of Phenomenological Constitution: A Self-Referential, Transcendental Approach to Conceptual Pathology* (Université de Paris, 1970; *Diss. Abs. Internatl.* No. 7905583).

and analysis, and as a phenomenon worthy of study in its own right, whether in cognitive science, artificial intelligence, general systems theory, the foundations of mathematics, epistemology, or other fields.

Both on the level of our factual dealings with the contingent world, and on the abstract level of theoretical necessities, the study of reflexivity has progressed within the last hundred years from a parlour curiosity to an indispensable and perhaps the most basic tool enabling us to gain an understanding both of ourselves as well as of systems whose dynamic, like our own, appears to be fundamentally self-regulating, and self-limiting.

· · ·

The essays contained in this volume consist of thirty-two papers on reflexivity, papers that form the classical basis for current research in this steadily growing area of study. They cover more than half of a century of work that was inspired, directly or indirectly, by the intellectual misgivings and confusion that followed the discovery of the semantical and set-theoretical paradoxes. These papers were originally published in numerous journals and volumes of conference proceedings, and have been brought together here for the first time, where they are printed in facsimile.

The individual essays were chosen for inclusion in this collection with three criteria in view: that each, when read in conjunction with others, should throw light on the evolution of thought about reflexivity; that each paper should, as we look back on the past sixty years of research, be recognizable as a basic contribution to current research; and that each article should point the interested reader on to other key contributions in the literature.

These essays, divided into families according to the varieties of reflexivity they examine, are of fundamental importance to an understanding and appreciation of the many-faceted, pervasive phenomenon of reflexivity.