

DIETER HENRICH: The Proof-Structure of Kant's Transcendental Deduction

Translated by KENLEY R. DOVE

The Review of Metaphysics, Vol. 22, No. 4 (June 1969) pp. 640–659.

THE PROOF-STRUCTURE OF KANT'S TRANSCENDENTAL DEDUCTION*
DIETER HENRICH
TRANSLATED BY KENLEY R. DOVE

The transcendental deduction of the categories is the very heart of the *Critique of Pure Reason*. It contains the two principal proofs of the book, the one demonstrating the possibility of a systematic knowledge of experience and the other the impossibility of knowledge beyond the limits of experience. Kant himself considered this theory completely new and extremely complicated; moreover he conceded that he had great difficulty in working out a satisfactory exposition of his theory. It is one of the two chapters which he rewrote completely for the second edition. Thus it is not surprising that this deduction has preoccupied interpreters more than any other text in the history of philosophy. In only thirty-five pages, which are easily separated from their context, Kant has formulated his most profound thoughts and presented the decisive foundation for his theory of knowledge. Whoever understands these pages possesses a key to the understanding and evaluation of the entire work. But Kant's text is so complex and elusive that it is difficult to follow the line of argument and to perceive within it the structure of a proof which could support the whole system of critical philosophy. In view of this it has been easy for Kant's critics to focus their attacks on the deduction. By the same token it has been just as easy for philosophers who wish to make use of Kant as testimony to their own position, to read their thoughts into his. Until now, however, no one has been able to offer an interpretation which agrees fully with Kant's text.

Hence, there is still controversy over which of the two versions of the deduction deserves priority and whether indeed any distinction between them can be maintained that would go beyond questions of presentation and involve the structure of the proof itself. Schopenhauer and Heidegger held that the first edition alone fully expresses Kant's unique philosophy, while Kant himself, as well as many other Kantians, have only seen a difference in the method of presentation.

In the following, an attempt will be made to settle this conflict which has persisted more than 150 years.¹ We shall advocate the thesis that only the second edition develops a tenable argument and that the argument in this version corresponds more adequately with the specific structure of Kant's thought as a whole, than does that of the first edition. This position contradicts the most important interpretations of Kant; moreover it proposes to reevaluate the meaning of his work and to guide its reception in a direction other than that of speculative Idealism, Neo-Kantianism, or Existential Philosophy.

I

We will treat first another controversy which, compared with the debate over the value of the two editions, is only of minor importance, yet which is relevant here insofar as it ultimately leads back

* Prof. Henrich's paper is reproduced here with his kind permission and with the translator acknowledgement (ed.).

¹ In this paper I shall discuss only the proof-structure of the Transcendental Deduction. An analysis of its premisses and the problems involved in the application of its conclusion will be given in another paper.

to this question and allows it to be answered: namely, the controversy concerning the structure of the proof in the second edition.

In this edition the conclusion of the deduction seems to be drawn twice in two completely different passages. It is the task of a transcendental deduction to demonstrate that the categories of our understanding are qualified to provide knowledge of appearances, as they are given to us in the unity of a synthesis of experience (B 123). The conclusion of section 20 reads: “Consequently, the manifold in a given intuition is necessarily subject to the categories” (B 143). This conclusion does not seem to differ from the result of section 26, according to which “the categories ... are ... valid a priori for all objects of experience”(B 161).

Thus one is tempted to see two proofs of the same proposition in the text of the second edition. That leads, however, into direct conflict with Kant’s unequivocal explication in section 21, which states that two arguments, rather than two proofs, are involved and that these together constitute the proof of the deduction. “Thus in the above proposition a beginning is made of a deduction of the pure concepts. . . .” “In what follows, ‘something further’ will be shown. . . .” “Only thus, by demonstration of the a priori validity of the categories in respect of all objects of our senses, will the purpose of the deduction be fully attained” (B 145). We can now formulate a criterion for a successful interpretation of the whole text of the deduction in this way: the interpretation must show that, contrary to the initial impression that the two conclusions merely define the same proposition, on the contrary, sections 20 and 26 offer two arguments with significantly different results, and that these together yield a single proof of the transcendental deduction. We shall call this task the problem of the two-steps-in-one-proof.

In previous commentaries this problem has been either pronounced insoluble or else passed over in silence.² The better commentaries claim that Kant’s assurance that his deduction presents two steps in one proof cannot be taken seriously, and that we are compelled to read the text as two distinct and complete proofs. Two proposals made on the basis of this double-proof theory merit our attention.³ We shall call them the interpretation according to Adickes/Paton⁴ and the interpretation according to Erdmann/de Vleeschauwer⁵ and shall examine them in that order.

1. In the preface to the first edition of the *Critique*, Kant himself distinguished an objective and a subjective side of the deduction (A XVI). The objective side makes the validity of the categories intelligible, the subjective investigates their relation to the cognitive faculties in us which must be presupposed if these categories are to be used. According to Kant one can also distinguish these two

² Cf., for instance, Norman Kemp Smith, *A Commentary to Kant’s Critique of Pure Reason* (London, 1918), p. 289; and A.C. Ewing, *A Short Commentary on Kant’s Critique of Pure Reason* (Chicago, 1967), p. 120.

³ In recent English publications on the *Critique* one does not find a discussion of this problem. Bennett and Strawson are writing from a point of view which does not focus on it. Wolff is interested almost exclusively in the first edition of the Deduction.

⁴ Erich Adickes, *Kants Kritik der reinen Vernunft* (Berlin, 1889), pp. 139–140; Herbert James Paton, *Kant’s Metaphysic of Experience* (London, 1936), vol. I, p. 501.

⁵ Benno Erdmann, *Kants Kriticismus in der ersten und in der zweiten Auflage der Kritik der reinen Vernunft* (Leipzig, 1878); Herman de Vleeschauwer, “La déduction transcendante dans l’œuvre de Kant,” in *Werken uitgegeven door de Faculteit der Wijsbegeerte en Letteren* (Ghent, 1937), vol. 3, pp. 24 et seq.

aspects as the demonstration *that* the categories have validity, and the demonstration *how* they attain validity. Adickes and Paton propose that this distinction be employed in order to understand the division of the deduction into two arguments: section 20 completes the proof of objective validity, section 26 demonstrates the subjective conditions of application.

This proposal has the advantage of being able to invoke in its support certain fundamental Kantian statements about the deduction—but there is no further evidence for it. For it is clear that the proposal cannot be applied to the structure of the second version of the deduction. In section 21 Kant clearly stated that the demonstration of the validity of the categories would be completed in section 26 (B 145). The title and conclusion of this section can be read in no other way. And the text itself contains no reflections about the interconnection of our cognitive faculties. The little word “how,” which can indicate the distinction between a psychological and an epistemological investigation, a subjective and an objective deduction, only appears incidentally. In this context, however, we shall see that it must be understood quite differently.

2. The proposal of Erdmann and de Vleeschauer likewise attempts to understand the second version of the transcendental deduction with the help of another observation of Kant’s—this time of a distinction made in the first version of the deduction. In two corresponding trains of thought, Kant here elaborates the relation between the categories, which can be developed from self-consciousness, and the given sensible representations. He distinguishes them as the demonstration “from above” and that “from below.” In this way he implies a hierarchy of cognitive faculties, the highest of which is the understanding and the lowest sensibility—extremes between which the faculty of imagination establishes a relation of possible coordination, and between which the two proofs move in opposite directions.

It seems quite natural to apply this distinction to the interpretation of the second edition. And indeed Erdmann and de Vleeschauer propose that section 20 be understood as a deduction “from above,” while section 26 is to be regarded as a deduction “from below.”

This proposal is in better agreement with the text of section 26, which has supplied the decisive arguments against the interpretation of Adickes and Paton. For Kant here actually proceeds from intuition, mentions the achievement of the faculty of imagination, and comes then to speak of the unity in the forms of intuition, which can be reached only through the categories and by virtue of the unification of the manifold in a consciousness (B 160). Nevertheless the two parts of the deduction remain unexplained by this proposal for the following reason: the structure of the first argument in section 20 can in no way be conceived as a deduction “from above”—and thus as a process which differs from the argument of section 26 insofar as its proof must be constructed in the opposite sequence. In section 20, just as in section 26, the manifold of a sensible intuition is mentioned first. Then it is shown that the manifold can assume the character of a unitary representation only if it is subject to the categories. Thus both arguments establish that a given intuition can become a unitary representation only when the intellectual functions of the understanding are applied to it. Now as to whether or not this argument can properly be understood as a deduction “from below”: the forms of

these proofs in no way make it possible to draw a meaningful distinction between the considerations of the two sections.

Hence the failure of the only proposed interpretations—not only because they depart from Kant’s assurance that there is *one* proof presented in *two* steps and attempt instead to find two distinct proofs, but also and primarily because their arguments can offer no useful explanation of the distinction between the two proofs.

We must search for another interpretation of the text. It should avoid both of these errors as far as possible and seek an understanding of the proof of the deduction that would require the two-steps-in-one-proof thesis. Moreover, it cannot derive support, as do the proposals just discussed, from Kant’s observations about the structure of the proof of the deduction, for they are valid only in the context of the first edition. Kant always allowed so many different trains of thought to influence him in formulating his central arguments that it is never possible to employ his comments unless he has explicitly related them to the passage of the text in question.

II

But now, from the propositions of sections 20, 21, and 26, we can develop a proposal which will solve the problem of the two-steps-in-one-proof. Its plausibility stems from the fact that it makes intelligible many peculiarities of the text which must be neglected by all other proposals.

Kant obviously attached importance to the fact that the result of the proof in section 20 contains a restriction: he established that intuitions are subject to the categories *insofar* as they, as intuitions, already possess unity (B 143). He indicates this restriction very clearly by writing the indefinite article in the expression “in an intuition” (*in Einer Anschauung*) with the first letter capitalized. Norman Kemp Smith, the translator, has misunderstood this hint.⁶ He believes that Kant wanted to imply that some single intuition was subject to the categories. This interpretation, however, leads to no meaningful emphasis in the course of the proof. Unlike English, in German the indefinite article (*ein*) and the word unity (*Einheit*) have the same root. This made it possible for Kant to express through the capital letter not the distinctness of any arbitrary intuition as opposed to others (*singularity*), but rather its inner *unity*.

The result of the proof in section 20 is therefore valid only for those intuitions *which already contain unity*. That is: wherever there is unity, there is a relation which can be thought according to the categories. This statement, however, does not yet clarify for us the *range within which* unitary intuitions can be found.

The restriction of the proof in section 20 is then discussed in that part of section 21 which makes reference to section 26. Here it is announced that the restriction just made in section 20 will be overcome in the paragraphs of section 26, i.e., the second part of the deduction will show that the categories are valid for *all* objects of our senses (B 161). And this is what actually takes place. The

⁶ Cf. p. 160 of his translation, which shows also that, as a commentator, he could not find sense in Kant’s text (cf. note 2 above).

deduction is carried out with the help of the following reasoning: wherever we find unity, this unity is itself made possible by the categories and determined in relation to them. In our representations of space and time, however, we have intuitions which contain unity and which at the same time include *everything* that can be present to our senses. For indeed the representations of space and time have their origin in the forms of our sensibility, outside of which no representations can be given to us. We can therefore be sure that every given manifold without exception is subject to the categories.

At this point the aim of the proof of the deduction has been attained, insofar as the deduction seeks to demonstrate the *unrestricted* validity of the categories for everything which can be meaningfully related to experience. Perceptions, which arise erratically and which cannot be repeated according to determinate rules, would not make intelligible a coherent and systematic knowledge of experience. The only conceivable result of a limited capacity for ordering our sense-data would be a diffuse and discontinuous sequence of perceptions.

It is certainly extraordinary to claim that our capacity for making conscious and thereby unifying our own sensuous representations should perhaps only be limited. However, its conceivability is an immediate result of the fundamental argument of the whole *Critique*. It is implied that our consciousness has the peculiarity of being “empty.” Everything of which we can become conscious must become accessible to us through media which do not immediately depend on this consciousness. According to Kant, it is for this reason that consciousness must be understood as an activity, thus always a *making-conscious* whose necessary inner unity causes us to give it the name “I.” But this activity always presupposes that something is present in the first place which is to be made conscious. Thus our consciousness can be found only together with a “passive,” receptive faculty, which is distinct and in certain respects opposed to the spontaneity of consciousness; it can encounter intuitions only as given “before all consciousness.” Kant reformulates the task of the transcendental deduction with reference to this very distinction: it must demonstrate that categories are capable of taking up something given into the unity of consciousness. “Appearances might very well be so constituted that the understanding should not find them to be in accordance with the conditions of its unity” (B 123). If that is possible, then it can also be asked whether such a disproportion between consciousness and givenness can be excluded for all or only for part of the given appearances. The difference between these two possibilities also defines the difference between the result of the proof of the first and that of the second step of the deduction.⁷

III

This question need not recur at every level in the analysis of the conditions of our knowledge. It could be that considerations are possible such as would establish rather quickly that the alternatives with which the transcendental deduction has to deal are not three-termed but rather only two-termed:

⁷ There was a thesis on the Transcendental Deduction by Friedrich Tenbruck (Marburg, 1944) never published, which came close to the conclusion of this section. Pietro Chiodi, *La Deduzione nell'Opera di Kant* (Torino, 1961), pp. 245 et seq. makes an attempt to bring the problem of the “how” (section 26) with that of the “that” (section 20) into a necessary connection so that together they build one chain of arguments. But one cannot distinguish the two sections on the basis of these two problems. And moreover, Chiodi’s account of Kant’s intention is highly formal and abstract and cannot be expressed in the language of the Transcendental Deduction itself.

that therefore either *no* sensuous representations or else *all* sensuous representations are capable of being determined by the categories. Anyone familiar with Kant's work will suspect that Kant had good reason to assert this. But this amounts to saying that Kant also had an alternative way of constructing the proof of the transcendental deduction, other than the one which he actually used in the second edition. For in this construction he takes into account the possibility of a merely partial ability of the understanding to establish unity in the sensible representations. He excluded it only because we do in fact have unitary representations of space and time and therefore can also unify all representations of sense.

Fortunately we can demonstrate that Kant himself was actually conscious of the fact that the transcendental deduction could also be constructed quite differently. His pupil Jakob Sigismund Beck undertook in the year 1793 to publish a selection from Kant's writings.⁸ On the title page he was able to announce that this was being done on Kant's own advice. Kant was interested in making available a competent commentary which could also be used for lectures. But when Beck published the third part of his selections in the year 1796, he considered it necessary to undertake a fundamental investigation in order to specify the standpoint from which Kant's *Critique* was actually to be evaluated. He had come to the opinion that the structure of the book promoted a false estimate of Kant's doctrine. Thus it would be necessary to begin with the productive activity of the understanding, in order to avoid the misunderstanding that Kant really wanted to speak of "given concepts" and of "objects which affect us." In Beck's opinion all this talk was only an accommodation to traditional doctrine and constituted preliminary concessions for the purpose of an introduction into the system. With this interpretation, Beck approached, somewhat belatedly, Fichte's philosophical conviction.

Naturally Kant could not bring himself to approve this. But since he was interested in Beck and in the effect of his writings, he was more willing to consider Beck's proposed alteration of the *Critique* than was his custom in comparable cases. In a letter to Beck's colleague Tieftrunk, he tried to show approximately what form the *Critique* might assume in an altered presentation.⁹ Thus we see that Kant himself at one time proposed an alternative to the transcendental deduction of the second edition.

It must begin with the doctrine of the categories as rules for the unity of a possible universal consciousness—corresponding to sections 16–18 of the second edition. Then it must demonstrate that intuitions a priori *are presupposed* in order that the categories can be applied at all to given sensuous intuitions. This becomes evident, when one considers that the categories can only be conceived as operators under which they can be applied. Without such a possibility of application an essential moment of their meaning is missing. The meaning of a priori concepts such as the categories can only be accessible a priori. But the only possibility of securing a meaning a priori for the categories is their application to a form of sensible intuition—the only a priori element which is conceivable in the domain of their application to sensible givenness. If there is no a priori intuition, then there is no

⁸ *Erläuternder Auszug aus Kants kritischen Schriften*, vol. 1 (Riga, 1793); vol. 3 (Riga, 1796).

⁹ *Kants gesammelte Schriften*, ed. Preußische Akademie der Wissenschaften, Briefwechsel, vol. 3, letter to Tieftrunk 11th Dec. 1797 and the first sketch of this letter in vol. 4, pp. 468 et seq.
Further evidence in vol. 5 of *Kants handschriftlicher Nachlass* in the same edition, reflections 6353 and 6358.

employment of the categories at all. Now the categories can only be applied a priori to intuition insofar as they grasp this form itself as a unitary representation. For categories are nothing else but forms of synthesizing into a unity. And by virtue of this, the application of the categories to all sensuous representations is also assured. For no sensible intuitions can be given *independently* of the forms of sensibility, which, in turn, are completely subordinate to the categories.

By reasoning in this way it is possible to maintain that the result Kant attains in section 26 on the basis of the mere fact of the givenness of unitary representations of space and time can be derived as a *necessary* condition of every employment of the categories. In accordance with this, the transcendental deduction can no longer be carried through in terms of those two steps of the proof which are characteristic of the second version.

In the same context, however, Kant also indicated the reasons for retaining the proof construction of the second edition: this proof makes use of the *synthetic method*, i.e., it proceeds on the basis of the fact that two doctrines of the *Critique* are initially developed independently of one another—the doctrine of the categories as functions of unity in self-consciousness and the doctrine of space and time as given representations. The second step of the proof according to the synthetic method has recourse to the results of the Aesthetic as to facts. If it were conducted according to analytical method, then the *necessity* of the forms of intuitions would first have to be justified. This would then be followed by an Aesthetic showing which forms we really have at our disposal. Only then could the deduction be completed. But Kant was of the opinion that this method “did not have the clarity and facility” characteristic of the synthetic method.¹⁰ And this is the reason which made him unable to consider Beck’s proposed construction as an improvement. Kant always had the tendency to make his theory convincing by virtue of its theoretical consequences and, as far as possible, to reduce analysis of its foundations to a minimum. He was intent upon changing the entire method of philosophical instruction and upon securing the convictions of his age against the dogmatists and against skepticism. He did not foresee that through this pressure for application he would disillusion the best speculative minds among his students and drive them to other paths.

IV

The papers documenting Kant’s reflections on the different methods for a transcendental deduction postdate the second edition of the *Critique* by almost ten years. Of course it is possible to show that all the ideas necessary for a deduction according to the analytical method had been already available to him when he published the first edition of the *Critique*. But this does not mean that he had in mind, as he composed the second edition, the advantages and disadvantages of a deduction according to one or the other method, and that he expressly chose the synthetic method on the basis of such a comparison. The text of the *Critique* provides no support for such an interpretation. Within the structure which Kant had already given his book, the advantages of a construction according to the synthetic method were in any case obvious. This construction allowed him to ground the two

¹⁰ Briefwechsel, vol. 4, p. 471.

fundamental positions of critical philosophy, the sensible a priori and the active role of the understanding in knowledge, separately—and then to unite them by means of a single argument.

But there were still other reasons which induced him to argue the proof of the deduction synthetically and to divide it accordingly into two steps. Besides the task of proving the objective validity of the categories, Kant also assigned to the deduction the task of making intelligible the possibility of relating the understanding to sensibility.¹¹ This task must not be confused with the other of which Kant speaks in the first preface to the *Critique*, when he distinguishes the subjective from the objective side of the deduction (A XVI). There he says that the subjective deduction is an investigation of those cognitive faculties upon which the possibility of a functional knowledge by means of the understanding rests. Such an investigation strives for more than the explanation of possibility. It seeks to elucidate the whole apparatus of knowledge, if only in a summary. The explanation of possibility proposes merely to remove a difficulty which arises out of the problem of critical philosophy itself: it assumes pure categories and then declares, however, that these categories are originally and essentially related to sensible intuition. Given the accepted idea of an a priori concept, this relation seems mysterious and inconceivable. And thus it must be shown that the meaning of an a priori concept can be so stipulated that it refers necessarily to intuition. And it must further be shown how one can represent the fact that the given intuition essentially depends on such concepts. This explanation of possibility can also be given in another form. It has then to show that categories as well as intuition cannot even be thought independently of their relation to one another. Moreover, the demonstration of the necessity of a relation between them can provide an answer to the problematical question concerning the possibility of their relations.

It is well known that Kant sought in the second edition to avoid the problems of the so-called subjective deduction. But that does not mean that he neglected the demand for an explanation of the possibility of relating the categories to intuitions. To be sure, Kant used the same words to distinguish between the two questions involved in the proof of the validity of the categories: the objective deduction is a proof *that* the intuitions are subject to the categories, while both the subjective deduction and the explanation of possibility are investigations of *how* they do this. But they are still two distinct investigations. Adickes and Paton have overlooked this distinction. For this reason they were obliged to consider the second step of the proof to be something which it clearly was not: a subjective deduction. At the same time, however, one may very well read the *whole* deduction as an explanation of the possibility of relating the categories to intuition.

Such an explanation, like the proof of validity, requires two steps of explanation. First it must be shown what the nature of a category actually is, given the fact that it is always at the same time related to a synthesis of intuition. And it must then be shown that such categories can exercise synthetic functions in intuition itself. These two parts of the explanation can be given *at the same time* with the two steps of the validity proof, according to which categories in general are valid without limitation. One cannot blame Kant for not separating the two investigations. For it is easily shown that the proof of the validity of the categories *must* enter into the explanation of the possibility of their

¹¹ B 159: “We have now to explain the possibility. . . .”

relation to intuition. At the only place where Kant separates the two investigations from one another, he was compelled to propose a proof of validity which fails to satisfy strict demands:¹² he has to *proceed* at this point *from* the *assumption* that we are in possession of synthetic a priori judgments concerning all objects of sensibility and that these judgments stand beyond all doubt in virtue of their employment in mathematical natural sciences. But this was the very presupposition which Hume called into question. And it is Kant's merit to have answered the radicalism of Hume's assault with a correspondingly radical founding of knowledge. For this reason, he ultimately did not incorporate into the *Critique* that form of a deduction which avoids every explanation of possibility. What we find in the second edition is a proof of the validity of the categories which is at one and the same time an explanation of the possibility of their relation to sensibility, a proof which avoids taking up the problems of an analysis of the cognitive faculties. And this is equally true of *both* steps of the deduction—not merely of its second part, which Adickes and Paton regard as a subjective deduction.

For an understanding of the text, both functions of the deduction must be borne in mind. But the proof of the validity of the categories always remains fundamental for a deduction. The second step of the proof, in section 26, overcomes the restriction still in effect in section 20. But considerable effort is required to recognize this second step. For it is not presented separately from the second part of the explanation of possibility, and Kant himself adds to the difficulty of understanding the argument when he declares, with great emphasis, that the unity in the representations of space and time can be none other than that which is thought through the categories.¹³ This statement, however, is only an application of the result of section 20 and of the conclusion of section 26. It contains neither a step of the proof nor a supplementary explanation of possibility. It is simply misleading to give an application so much stress that the actual conclusion of an argument is lost sight of.

V

But even with all these considerations it has still not been sufficiently clarified why Kant never presented the actual structure of his proof in a clearer manner or never even indicated it unequivocally. We have been able to reconstruct it only from a text which constantly involves other elements and at times gives them undue emphasis. A further reason can be given for this circumstance which leads into a fundamental consideration about the second version of the deduction and its relation to the first. Already in section 16 Kant seems to suggest that he has completed the proof of the deduction that all sensible representations are subject to the categories. Kant's argument at this point makes use of an analysis of the meaning expressed when I say that a representation is my representation. The unity of apperception is the original definition of the meaning of "belongs to me." For this unity is indicated by the expressions "I" or "I think." All representations are therefore mine insofar as they are bound together in the unity of self-consciousness. But Kant now extends the meaning of "mine" further, namely, to include all representations which *can* be united in self-consciousness. There are good grounds for this extension. For we do actually call representations ours in virtue of the fact that

¹² Cf. the note to the introduction to Kant's *Metaphysical Foundations of Science*. Like the *Prolegomena* it starts from science as an indubitable fact, which is, according to the *Prolegomena*, legitimate only within an introduction into the *Critique*. The differences between the *Prolegomena* and the note may be ignored here.

¹³ B 161: "This synthetic unity can be no other than. . . ."

we can become conscious that they belong to us. And there is no better criticism for the association of representations with a consciousness than the fact that it can experience them as its own representations.

This extension is equivocal nevertheless. For it leads easily to the more extensive thesis that all representations which arise in the sensibility of a being are to that degree also already potentially *conscious* representations. Precisely because every consciousness presupposes, according to Kant, a sensibility, one is tempted to describe this sensibility itself as “his” sensibility and then further to assume that all given sensible representations must also to that extent be “his.” This would mean that they are to be defined as possible contents of his consciousness. And, by proceeding in this way, the transcendental deduction would be completed as a result of a simple semantic analysis of how the word “mine” is used. For if all given representations are “mine” in the sense indicated, then that means precisely that they can be taken up into the unity of consciousness in accordance with the categories.

That would be an astonishingly simple solution to the problem which Kant had announced as the *deepest* in the whole transcendental philosophy. Actually it rests, as is obvious, on a shift of meaning in the expression “mine.” Sensibility is distinct from self-consciousness. And if something belongs to me only if I can take it up into consciousness, then as long as it is only *available* to be taken up into consciousness, it is not at all “mine”; but only “in relation to me.” It makes no difference whether this relation is public or private. If the limits of my consciousness’ capacity to take something up precludes its entering my consciousness, it would also never become “mine” in the strict sense of the word.

Kant was certainly aware of this distinction. At an important point he refrains from saying that there could be no intuition at all which would remain inaccessible to consciousness. He affirms only that this representation would then be “nothing for me” (B 132). But it must also be noted that Kant did not seriously assume that there could be such representations—and not merely in virtue of the proof of the deduction, whether it be construed according to the analytical or the synthetic method. He did not see with sufficient clarity the homonymy of the word “mine” in talk about “my” representations. He was thus able to assert propositions which anticipate the result of the proof of the deduction and at the same time make the deduction dependent on the mere semantic analysis of the word “mine”: “I am conscious of the self as identical with respect to the manifold of representations that are given to me in an intuition, because I call them one and all *my* representations, and so apprehend them as constituting *one* intuition. This amounts to saying that I am conscious a priori of a necessary synthesis of representations—to be entitled the original synthetic unity of apperception—under which all representations that are given to me must stand ...” (B 135). In the first edition Kant made use of an even poorer argument in order to make the same conclusion convincing (A 120). In the second edition one can clearly recognize that he could not free himself from such arguments, but also that he does not rely upon them confidently. And for this reason he never expressly stated that the deduction already takes place in section 16. Instead he assures us that it *requires* both of the steps which we have analyzed (B 145). And they make no use at all of the semantic analysis of the word “mine.”

One could show that it was no accident that Kant was upset by the ambiguity of this word. The difficulty inevitably occurs if one takes his doctrine of consciousness as a point of departure. Its distinctive features, however, cannot be examined here.

For the present discussion it is more important to see that this difficulty led to entirely different consequences in the two editions of the *Critique*. This difficulty is at least partly responsible for giving the second edition deduction an external form which is confusing and misleading. If we take the necessary pains, we can, nevertheless find an argument there which is convincing in the context of Kant's philosophy. In the first version, however, there is *no* proof which can dispense with the misleading argument from the double meaning of "mine." One can easily see this if one tries to rediscover in the first edition that thought which is indispensable for the second edition's division of the proof into two steps: the idea that the representations of space and time themselves presuppose a synthesis which includes everything that is given in them. To be sure, Kant took over this idea from the first edition, but at the same time he gave it an entirely different function. There it occurred only in the *introduction* to the proof (A 99, 101–2, 107). It seeks to clarify the fact that our cognitive process consists in something more than empirical powers and principles of combination which could only be investigated by association-psychology. An a priori synthesis is also involved. Evidence of this is given in our representations of space and time, which cannot be understood without a priori synthesis. And this kind of synthesis leads to the philosophical question concerning the conditions of its possibility. By means of this argument Kant thus introduces in the first edition those principles in terms of which a transcendental deduction is to be given. In the second edition it has become an essential *part of the deduction* itself.

Thus there is a substantial difference between the proofs of the first and the second editions, and not merely in the manner of presentation, as Kant himself believed. We can understand why he himself was *unable* to see this distinction. For even in the second edition he did not yet altogether give up the inadequate argument that had been *absolutely* indispensable for the proof of the first edition. But as long as incompatible proof-strategies still continued to affect Kant's conception of the argument, he was not able to give an unequivocal presentation of the actual structure of the proof and to distinguish it clearly from the many related intentions which a transcendental deduction has to take into account at the same time. A careful stylistic analysis of the second edition reveals that Kant has modified his language in section 20 and that thereby he reaches for the first time the paths of the deduction which were to offer a cogent argument.

VI

We have noted previously that the proof of the second edition has the advantage over a possible analytical line of argument that it is better suited to the structure of the *Critique*. Now it has become apparent that, in comparison with the proof of the first edition, it also has the more significant advantage of being a formally correct proof. But these are not the only reasons for believing that Kant's thinking is more appropriately expressed in the second version of the deduction. An analysis of the proofs yields a far more general conclusion.

Now if it were only the structure of the book which recommends the proof of the second edition, one could, with Beck, suggest an alternative external shape for the *Critique* and thus a change in the form of its central argument. In point of fact, however, the second deduction is in complete agreement not only with the structure of the book but also with Kant's fundamental philosophical idea about the methodology of a philosophical system.¹⁴

Kant based this system on the unifying principle of the unity of self-consciousness. All its propositions must be deducible from this principle. But the method of this derivation does not consist in developing analytical implications from the concept of self-consciousness. Instead it specifies the presuppositions of the *possibility of the existence* of self-consciousness. By using this method, one can come to a knowledge of conditions which, although they are not already given in the structure of self-consciousness itself, must precisely *in virtue* of this structure be presupposed if a self-consciousness is to become actual.

This manner of argument is distinctively Kantian. It combines two propositions which may be regarded as the two formal tenets of the Kantian system: there must be a unifying principle in terms of which knowledge can be understood; yet this principle must not entail a monistic exclusion of all other principles; it must take into account the discovery of the essential difference in the roots of our knowledge and make possible a *raisonnement* which presupposes their underivable synthesis.

This method of argumentation is just as distinct from the faculty-psychology of empiricism as it is from those doctrines of the universal implications of the Ego which transformed Kant's position into that of speculative idealism. Empiricism lacked any principle of unity. The form of idealism which was historically so influential had no concept of an essential unity of originally *distinct* elements. Kant's transcendental deduction—*but only in its second version*—contains such a concept, although it is not fully developed. It proceeds, albeit with difficulty, according to a method of proof which is oriented by this concept.

If one succeeds in reaching a clear understanding of this method, one possesses the key to an understanding of Kant which is independent of his specific formulations. But what is more important, one will be able to understand those problems which remained insoluble for Kant as well as for his successors: the structure of consciousness, of moral knowledge, and of temporal experience.

It was only very late in his philosophical development that Kant worked out a balance between the two elements of such a method. First he realized the necessity of assuming distinct principles of knowledge whose interrelation is necessary. Later he discovered the unitary principle in terms of which such relations must be interpreted. Under the influence of this discovery, however, he maintained for more than a decade plans for a deduction which transcended the capacity of the unifying principle and which brought him into contradiction with his earlier discovery. Thus, for example, one can show that his *moral* philosophy was freed from more extensive claims of a deduction from self-consciousness

¹⁴ D. Henrich, "Zu Kants Begriff der Philosophie," in *Kritik und Metaphysik, Festschrift für Heinz Heimsoeth* (Berlin, 1966), pp. 40 et seq.

and assumed its ultimate form only shortly before the appearance of the *Critique of Practical Reason*.¹⁵ The change in the conception of a transcendental deduction corresponds fully to this development. And this correspondence is certainly not the weakest confirmation of the proposed interpretation.

Kant himself did not reach a clear understanding of the nature of these changes. And he withheld from his readers the clarity which he actually did possess for strategic reasons. Thus one cannot blame his successors if, unable to discover the coherence of his texts, they began to seek the spirit of critical philosophy in that conception of the nature of a philosophic system which Kant himself had just recently dismissed. In any case they were right insofar as the peculiar content of critical philosophy is only disclosed by autonomous philosophical effort. This task is still uncompleted today. But the solution of the enigma involved in an interpretation of the transcendental deduction shows where this effort has to be directed.

University of Heidelberg.

¹⁵ This is shown in various articles of mine, among them: “Der Begriff der sittlichen Einsicht und Kants Lehre vom Faktum der Vernunft” in: *Die Gegenwart der Griechen im neueren Denken* (Tübingen, 1960), pp. 77 et seq.; and “Das Problem der Grundlegung der Ethik bei Kant und im spekulativen Idealismus” in *Sein und Ethos* (Mainz, 1963), pp. 350 et seq.