Sidgwick's Legacy? Russell and Moore on Meaning and Philosophical Inquiry

Sébastien Gandon

James Levine has recently argued (1998, 2009, 2016) that there is a tension between Russell's Moorean semantical framework and Russell's Peano-inspired analytical practice. According to Levine, this discrepancy runs deep in Russell's thought from 1900 to 1918, and underlies many of the doctrinal changes occurring during this period. In this paper, I suggest that, contrary to what Levine claims, there is no incompatibility between Moore's theory of meaning and the idea of informative conceptual analysis. I show this by relating Moore's view of meaning to his Sidgwick-inspired criticism of the so-called naturalistic fallacy. I maintain that Moore's semantical framework has a methodological intent: following Sidgwick, Moore wants to block any attempt to justify ethical principles through setting *ad hoc* conditions on the meaning of the terms involved. Thus, far from grounding philosophical knowledge on subjective intuitions, as Levine suggests, Moore's framework would provide us with the means to make room for a discursive and dialectic conception of philosophical inquiry.

Sidgwick's Legacy? Russell and Moore on Meaning and Philosophical Inquiry

Sébastien Gandon

1. Introduction

In the preface of *The Principles of Mathematics*, Russell wrote:

> On fundamental questions of philosophy, my position, in all its chief features, is derived from Mr G. E. Moore. . . . Before learning . . . from him, I found myself completely unable to construct any philosophy of arithmetic, whereas [the] acceptance [of his views] brought about an immediate liberation from a large number of difficulties which I believe to be otherwise insuperable. [Moore's] doctrines . . . are, in my opinion, quite indispensable to any . . . philosophy of mathematics. (Russell 1903a, xviii)

One page later, he added:

> Many words will be found, in the course of discussion, to be defined in senses apparently departing widely from common usage. Such departures, I must ask the reader to believe, are never wanton, but have been made with great reluctance. . . . But in many instances of such apparent departure from usage, it may be doubted whether more has been done than to give precision to a notion which had hitherto been more or less vague. (Russell 1903a, xviii–xix)

In a series of important articles (1998, 2009, 2016), James Levine maintains that these two passages are at odds with one another. The former of the two shows the influence of Moore; the latter, in a more subtle way, underlines that of Peano. The paradigmatic example of analysis is indeed the famous one of cardinal numbers in terms of classes, developed in the context of Peano's "logistic". Levine's thesis is that Moore's philosophy, and more

particularly his theory of meaning and understanding, is incompatible with Peano's inherited analytical practice.

Let me explain. Moore, during the years 1898 to 1903, views the understanding of an expression (a name or a statement) as a direct cognitive contact between a mind and the entity (the meaning) designated by the expression. This view does not leave any room for the idea of a partial or incomplete understanding: either the mind relates to the appropriate entity and, in this case, the understanding is complete, or the mind is not in acquaintance with the designated entity, which means that nothing is understood. In other words, Moore's theory of meaning rules out the possibility of an informative analysis, i.e., an analysis which does not only come to a simple repetition (or to an explication) of a meaning that is already given. Yet many Russellian analyses in the *Principles* are informative—the analysis of cardinal numbers in terms of class, for a start. There is thus a deep tension in *The Principles* between a theory of understanding based on cognitive contact and Russell's practice of analysis. According to Levine, this tension is at the heart of Russell's thought during the period that stretches from when he met Peano in 1900 to the beginning of the 1920s. Russell's many evolutions during this interval are seen by Levine as attempts to adjust the semantical framework to the analytical methods.[1]

Such an interpretation is questionable. The claim on which it is based seems to be fragile. The aim of this article is to establish that an alternative interpretation of the methodological implications of Moore's theory is possible. More precisely, I would like to show that when Moore's considerations about meaning are connected with his discussion of the naturalistic fallacy and Sidgwick's utilitarianism, a new picture of Moore's views emerges.

It is important not to be mistaken about the scope of my criticism. My intent is not to directly confront Levine's interpreta-

---

[1]On the issue of partial understanding and its importance, see also Burge (1990) and our conclusion.

tion of Russell. I only maintain that one important ingredient on which Levine's reading is based (his understanding of the methodological implications of Moore's semantical framework) should be qualified. I do think that this has some implications on the way one should interpret Russell, and I give some arguments for that at the end of the paper. But I don't claim to have proved this. The broad thrust of the article is focused on Moore, and on Levine's reading of Moore.

Section 2 presents the main lines of Levine's interpretation. In Section 3, I present the contents of Sections 10–14 of *Principia Ethica*, in which Moore criticizes an argument, called the naturalistic fallacy, aiming to justify principles by resorting to stipulative definition. This argument, taken up again in Part III of Moore (1903a), is in fact, as I will show in Section 4, a generalisation of a criticism which Sidgwick addressed to the first utilitarians, Bentham and Mill. In Section 5, I show that, when anchored in this context, Moore's theory of meaning does not have the methodological consequences that Levine attributes to it. In Section 6, I briefly come back to Russell, and give my own explanation of the difference between Moore and Russell concerning definitions.

2. Levine's Interpretation

According to Levine (2016, 168), Moore's theory of meaning is grounded on two pillars. The first is the principle according to which

Augustinian Principle (Aug): The meaning of a word is an entity (simple or complex) corresponding to that word; the meaning of a sentence is a proposition—a complex entity—whose constituents are the meanings of the words in that sentence.

The second pillar is the principle according to which

Principle of Acquaintance (PoA): Apprehending a proposition requires being acquainted with each of its constituents.

(PoA), as formulated here, does not indicate whether the apprehension of a complex entity (for example a proposition) requires, as well as acquaintance with simple constituents, an extra-acquaintance with the complex entity itself. The question is difficult, but can be left aside.[2] Note also that the meanings in question in (Aug) and (PoA) are not mental states; they are entities that exist outside the mind. Thus, even if Moore does not use the term "acquaintance" that Russell will popularise, Levine is right to assimilate the relation of mind to meaning in Moore to Russellian acquaintance.[3]

Levine argues that Moore's semantical framework has two major methodological consequences, which concern both conceptual analysis and philosophical argumentation. Levine first explains that (Aug) and (PoA) lead us to disqualify analysis as a source of knowledge:

Accepting (Aug) along with (PoA) places a severe constraint on what is involved in analyzing the proposition expressed by a given sentence. For by (Aug) together with (PoA), given a sentence *S*, its meaning is a proposition *P*, and understanding *S* requires being acquainted with each constituent of *P*. In that case, analyzing *P*— a task that involves identifying its ultimate constituents—involves making explicit the entities with which anyone who understands

---

[2]The important thing for us is that an acquaintance with all the simple terms is required to understand a proposition. On the question of whether acquaintance with the simple constituents is enough in order to have an acquaintance with the complex that they make up, see Russell (1992, 119–28).

[3]In Russell as in Moore, the mind has a direct cognitive relation with a non-mental entity. This entity may be sensitive, as well as non-sensitive, purely intellectual—as in the case with purely logical objects in Russell. The non-mental character of sense data is at the centre of the refutation of idealism in Moore (1903b, 42): "There is no need to wonder how to 'get out of the circle of our own ideas and sensations'. The mere fact of having a sensation brings us out of the circle".

S is acquainted. That is, it will involve making explicit what is already "present to the mind" of one who, prior to analysis, understands *S* (or, indeed, any sentence expressing *P*). (Levine 2016, 171)

Moore's theory of meaning makes understanding an all or nothing case: either the mind has a contact with the designated entity, or it has acquaintance with another one (or with nothing at all). In the second case, because a wrong thing (or nothing) is understood, no analysis is possible. However, in the first case, the analysis is, in a way, already carried out before having been initiated. The analysis of a proposition can only be, in this perspective, considered as a trivial operation, consisting in repeating (or at best, making explicit) a content that has already been grasped. In other words, the adoption of (Aug) and (PoA) brings us to the following thesis:

(M1) Conceptual analysis is not informative and does not bring about any new knowledge.[4]

Levine extends this first development by a second one which concerns the justification of primitive propositions. According to him, Moore and Russell hold "that the characteristic philosophical activity consists of 'intuitively apprehending' the meaning of indefinable terms . . . and 'intuitively apprehending' the truth of indemonstrable propositions" (2009, 30). This intuitionism leads to contrasting philosophical method, based on direct intuition, with scientific method, based on rational argumentation. Levine thus maintains that:

Since it is ultimately up to each individual to intuit what he or she intuits, philosophy becomes "in a certain impersonal sense controversial". For if you fail to intuit what I have intuited, there is no rationally compelling argument by which I can convince you that what I have intuited is really there; the most I can try to do is to "cause" you "to perceive" what I have "perceived", so that what is involved here is not rational argument but rather "exhortation". If my "exhortation" fails, we are left—"in a certain impersonal sense"—with a dispute that cannot be rationally adjudicated. (Levine 2009, 32)

For Moore and Russell, philosophical research would thus neglect rational argumentation ("*does not essentially involve argument or reason-giving*", 32) to become a simple matter of directly perceiving what is true within each subjective mind. In other words, (Aug) and (PoA) lead to the following thesis:

(M2) The justification of first principles is not founded on argumentation, but on immediate and subjective intuition.

Having come this far, one can understand how Moore's theory of meaning rules out the Peano's inspired analyses one encounters in *The Principles*.

It is not easy to find a definition of analysis that covers all the various cases dealt with in Russell (1903a). But whatever the instance considered, this much is clear: Russell's Peano-inspired analysis[5] of a given concept does not amount at making explicit what is, in some sense, already "present to the mind" of anyone who understands the notion, prior to analysis. Let me take the emblematic example of the definition of cardinal numbers in terms of equivalence classes:

---

[4]The thesis is close to the one that will lead Moore to formulate of his paradox of analysis (1942, 665f.). Thus, Levine (2016, 171): "in applying a Moorean conception of analysis, the early Moore and Russell are often not attempting to provide apparently 'informative' analyses of familiar concepts, but rather argue against apparently 'informative' analyses provided by others, on the basis that such analyses conflict with what is 'present to the mind' of one who understands the words in question."

[5]By Peanian or by Peano-inspired analysis, I intend the series of analyses displayed in Russell (1903a)—that is, at least, the analysis of finite numbers, of order, of progression (ω-sequence), of continuity, of infinity, of quantity, of projective space, and of distance. For attempts to characterize Russell's 1903 analyses in general terms, see Hylton (1990) and Levine (2007); for an account that does not assume any such general characterizations, see Gandon (2012).

(Num) The cardinal number of $\alpha = \{\omega : \omega$ is similar to $\alpha\}$.

The identity (Num) constitutes, if it is correct, new knowledge, an informative analysis of the notion of number: we do learn something we didn't know before when we learn that a number is a class of similar classes. And this goes against (M1). Moreover, the legitimacy of (Num), far from relying on an immediate intuition, is based on theoretical advances, particularly on the works of Cantor, Dedekind and Peano. And this goes against (M2).

Moore's reaction to the Russellian definition of cardinal numbers is instructive in this respect. In the unpublished manuscript in which Moore commented on the *Principles*, he said:

> The definition in logical terms of number 'one' is by no means simple . . . It is not plain that what we think to be true of the penny, when we think it is but one, is no less than that it is a member of the class of [one-membered] classes . . . : it is not plain that this is a correct analysis of what we think. That it is equivalent to what we think, in the sense that anything whatever which has the property which we mean by 'one' is also a member of this class of classes, and that anything whatever which is a member of this class of classes also has the property which we mean by 'one', there is, indeed, no doubt whatever. (Moore 1905–06, 8–9)

Although, mathematically speaking, there is an equivalence between the meaning of "one" as it is grasped by an ordinary person in ordinary circumstances, and the class considered by Russell, it nevertheless remains that both notions have different meanings and therefore that they must not be identified with each other. It is therefore because he wants to remain faithful to his semantical framework that Moore rejects (Num). The same kind of argument based on introspective intuition may be mobilised to disqualify Russell's other famous 1903 analyses (those of order, limits, continuity, infinity, space, etc.). Levine seems right: there seems indeed to be a tension between the Peano-inspired analyses that made Russell's name, and the con-

sequences (M1) and (M2) of Moore's theory of meaning.[6]

Before discussing Levine's interpretation, one point is worth going over. The (M2) case is more complicated than the (M1) case. (M2) is about truth and propositions, and neither (Aug) nor (PoA) deal specifically with truth and propositions. Something is then lacking in the account given by Levine. In order to deduce (M2) from Moore's theory of meaning, another principle, dealing specifically with the propositional case, is required. The following "Principle of Evidence" is an attempt to fill the gap:

Principle of Evidence (PoE): First philosophical principles cannot be proved, their truth can only be intuitively apprehended.

From (PoE), one can conclude (M2). Levine does seem to attribute (PoE) to Moore, when he claims that, for Moore, the truth of indemonstrable propositions is intuitively apprehended (2009, 30). I take the formulation of (PoE) thus as a way of spelling out a feature left implicit in Levine's account. But as it will soon become clear, I think that Levine's failure to make explicit the distinction between the propositional and the non-propositional cases has important consequences on the interpretation of Moore's methodological views.

The following schema (S1), in which the first line contains the principles of the theory of meaning and the second shows the methodological and metaphysical consequences of these principles, illustrates the relationship between Moore's different theses:

(S1)

(Aug) + (PoA)	(PoE)	Theory of meaning
↓	↓	
(M1)	(M2)	Philosophical methodology
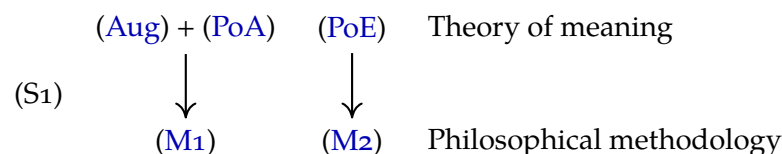
The arrows represent implication relations.

I have no doubt that Moore subscribes to (Aug), (PoA) and (PoE). I also grant that Moore's semantic principles have important consequences for philosophical methodology. However, I do contest that (M1) and (M2) are consequences of (Aug), (PoA), and (PoE). But before challenging Levine's interpretation, I need to look at the immediate context of Moore's theory: the critical discussion of the so-called "naturalistic fallacy" at the beginning of *Principia Ethica*.

3. The Naturalistic Fallacy and the Theory of Meaning

To show that Moore subscribes to (Aug) and (PoA), Levine refers to Section 13 of the *Principia Ethica* (Moore 1903a; see Levine 2009, 16, 2016, 169, 171–72; Levine also refers to Moore 1903b) which belongs to a development (§§10–14) dedicated to the discussion of the naturalistic fallacy. I will here analyse the nature of the fallacy and the role that (Aug) and (PoA) play in its criticism.

As Baldwin showed (1990, 69ff.), the naturalistic fallacy is more of a gathering of distinct stances rather than an unequivocal thesis. Two relatively coherent elements can be distinguished within this disjointed bundle. The fallacy is first of all described as a (flawed) conception of the nature and the role of definitions. This methodological line is particularly developed in Parts I and III of *Principia Ethica*. The fallacy is secondly presented as a (mistaken) view about ethics: it consists of considering that Good is a natural quality and, as a consequence, of confusing norms and facts, "ought" and "is". This is the line that is unfolded in Parts I, II and IV of the book. There is of course no immediate connection between the two ideas. Moore however groups these two positions together, without ever distinguishing them clearly. It was only in 1932, when he came back to the work of his youth, that Moore acknowledged that the fallacy is a combination of

different stances.[7] As Baldwin notes, this lack of clarity on such an important subject makes it difficult to make sense of Moore's 1903 ethical project (Baldwin 1990, 66–73, 86–100).

Luckily for us, that is not the task we are confronting. In Sections 10–14, where the fallacy is presented for the first time, it is only the methodological dimension that is put forth. Even if they are not free from difficulties (Baldwin 1990, 86–89), Moore's statements in those sections are relatively coherent. In substance, Moore relates the naturalistic fallacy to the confusion between two ways of presenting a meaning: *via* a genuine definition, which decomposes the targeted meaning into its constituents; and *via* what I would call (for lack of a better term) an univocal characterization, which picks out the meaning by referring to a property[8] that it is the only one to possess.

For Moore, meanings are either simple or complex, and a complex meaning is composed of meanings that are themselves either simple or complex. Defining a meaning is breaking down a complex meaning into its simple constituents: "the most important sense of 'definition' is that in which a definition states what are the parts which invariably compose a certain whole" (1903a, 9). For example, the complex meaning *horse*,[9] designated by the word "horse", can be broken down into different elements which are, according to Moore, parts of the horse. When all the parts in the definiens are simple, then the analysis stops and is complete.

All this is quite clear and simple. What makes the situation complicated is that meanings (either simple or complex) can be presented in another way. It could be the case that a meaning *E* is the only one to fall under a certain concept *C* (no other

---

[7]See Baldwin (1990, 71, 317), who refers in particular to a draft for a revised preface to *Principia Ethica* written around 1920.

[8]Usually, the property used is a relational property, i.e., the property of having such and such relation to such and such entities.

[9]To indicate that I am talking about meaning of the sign "*A*", I will add stars, as in: *A*.

entities satisfy the property *C*). When this happens, it is possible to characterize *E* by using the expression "the C",[10] since *E* is the only one entity to which the phrase "the C" applies. Thus, Moore explains:

> Consider yellow, for example. We may try to define it, by describing its physical equivalent; we may state what kind of light-vibrations must stimulate the normal eye, in order that we may perceive it. But a moment's reflection is sufficient to shew that those light-vibrations are not themselves what we mean by yellow. *They* are not what we perceive. . . . The most we can be entitled to say of those vibrations is that they are what corresponds in space to the yellow which we actually perceive. (Moore 1903a, 10)

The phrase "what is caused by stimulating the normal eye by such and such luminous vibration" does apply to the colour yellow, and to no other colours. In this sense, it univocally characterizes the entity *yellow*. But how to articulate the genuine definition with this kind of characterization?

For Moore, definition and characterization are two completely different things. The entity *E* is designated by "E"; the entity *the C* is designated by "the C"; and *E* and *the C* are not the same. In particular, the latter cannot be taken as a decomposition, or analysis, of the former. For instance, even though *yellow* (and no other colour) can be characterized as "what is caused by stimulating the normal eye by such luminous vibration", it is not the case that *yellow* is composed of *cause*, *normal eye*, *luminous vibration*.

This distinction between definition and characterization is crucial, since it lies at the core of Moore's criticism of the naturalistic fallacy. But before elaborating on this, I need to make three remarks. First, the issue concerning the difference between definition and characterization is independent from the question of knowing whether the targeted meaning is simple or complex.[11] When the meaning is simple and the term is not definable, it can still be, in many ways, univocally characterized. In a case where the meaning is complex, the decomposition of the meaning in a definition must still be carefully distinguished from a univocal characterization. Second, the distinction between definition and characterization is in line with Moore's endorsement of (Aug) and (PoA). Indeed, what allows the speaker to maintain that the meaning of "E" is different from the meaning of "the C" is that his direct acquaintances (*via* (PoA)) with what the two names univocally designate (*via* (Aug)) are different. Thus, the two semantical principles (Aug) and (PoA) provide the speaker with a principled reason to distinguish definition from characterization. I will come back to this soon. Third, in Chapter 5 of *The Principles of Mathematics*, Russell elaborates a theoretical framework[12] that appears to be close to what Moore develops here. Russell's *the*-denoting concept[13] seems to correspond to what I have just called univocal characterization. Indeed, Russell stresses that a denoting concept *the C* that applies to a single entity *E* must nevertheless be carefully distinguished from the single entity it denotes. Exactly like a univocal characterization in Moore, a *the*-denoting concept applies to one single entity, and yet cannot be considered as a genuine decomposition of it. The connection between Russell's theory of denoting concepts and Moore's analysis is, however, more complicated than it appears at first. I will come back to this issue in Section 6.

---

[10]The occurrence of "the" brings Moore's analysis close to Russell's theory of denoting concepts. On this, see below. Note however that one does not find any discussion of the role of the definite article in Moore (1903a) and that the occurrence of "the" is not a necessary condition for univocal characterization.

[11]In (1903a, §13), Moore elaborates an argument aiming to show that the meaning of "Good" cannot be complex. I will leave this point aside. About the link between the naturalistic fallacy and the simplicity of *Good*, see Baldwin (1990, 66–73).

[12]More on this in Section 6 below.

[13]Russell (1903a, 62): "The word *the*, in the singular, is correctly employed only in relation to a class-concept of which there is only one instance. We speak of *the* King, *the* prime minister, and so on (understanding *at the present time*); and in such a case there is a method of denoting one single definite term by means of a concept."

Now, as I said before, the naturalistic fallacy consists in confusing definition and characterization. The passage about yellow continues as follows:

> Yet a mistake of this simple kind has commonly been made about 'good'. It may be true that all things which are good are also something else, just as it is true that all things which are yellow produce a certain kind of vibration in the light . . . But far too many philosophers have thought that when they named those other properties they were actually defining good; that these properties, in fact, were not 'other', but absolutely and entirely the same with goodness. This view I propose to call the 'naturalistic fallacy'. (Moore 1903a, 10)

The naturalistic fallacy is thus the view that when one univocally characterizes a certain entity *E* in a certain way, one is actually defining it. "Far too many philosophers", explains Moore, have thought they could replace the meaning of "good" ("E") with a univocal characterization referring to it ("the C"), and this belief constitutes the naturalistic fallacy. How is this confusion important for present purposes?

Section 11 gives us some clues. The key point is the emphasis put on the dialectical context of a disagreement between two philosophers concerning the truth of an ethical principle. Let us imagine that a philosopher S_1 maintains (thesis T_1) that a certain concept *A* (for example, Good) is identical with *P_1* (for example, what is pleasant), and that a second philosopher S_2, disagreeing with S_1, maintains (thesis T_2) that *A* is *P_2* (for example, what is desirable)—with P_2 different from P_1. How can S_1 convince S_2 of the truth of T_1? One possible strategy for S_1 would be, explains Moore, to argue that, by definition, the meaning of "A" is *P_1*. S_1 would not only argue that *P_1* is a property of *A*, nor that *P_1* is a property that *A* is the only one to have—but that the meaning of "A" is *P_1*. According to this hypothesis, the truth of T_1 would derive from the meaning of "A": T_1 would be an analytic proposition. This is the crucial point. The strength of the naturalistic fallacy strategy comes from the fact that T_1 seems to be proved: S_2 can only object to T_1 by dismissing what is presented as an analysis of *A*, i.e., by admitting that he does not know what "A" means.

The problem with this strategy is, however, that it does not erase the disagreement. S_2 can indeed go ahead and maintain that the meaning of "A" is *P_2*, and that the thesis T_2 is an analytic truth, resulting from the definition of *A*. So how can the dispute be settled? In the next part of Section 11, Moore imagines two possible scenarios. S_1 can seek to prove that, with *P_1* different from *P_2*, *A* cannot be *P_2*. But by doing so he presupposes what is to be demonstrated, namely that *A* is *P_1*:

> The position is like this. One man says a triangle is a circle: another replies, 'A triangle is a straight line, and I will prove to you that I am right: *for*' (this is the only argument) 'a straight line is not a circle.' 'That is quite true,' the other may reply; 'but nevertheless a triangle is a circle, and you have said nothing whatever to prove the contrary. What is proved is that one of us is wrong, for we agree that a triangle cannot be both a straight line and a circle: but which is wrong, there can be no earthly means of proving . . . .'—Well, that is one alternative which any naturalistic Ethics has to face; if good is *defined* as something else, then it is impossible either to prove that any other definition is wrong or even to deny such definition. (Moore 1903a, 11)

The other possibility is to transform the question into a purely verbal discussion, concerning the rules that govern the use of expressions "A", "P_1" and "P_2". S_1 could argue that when he states "A is P_1", he means that "most people" (1903a, 11) have used the word "A" as a substitute for "P_1". But for Moore this is yet again changing the subject. From a question about *A*, we are being led to a question about the use of the word "A" (on this point, see also Baldwin 1990, 197–202):

> The other alternative will scarcely be more welcome. It is that the discussion is after all a verbal one. When *A* says 'Good means pleasant' and *B* says 'Good means desired,' they may merely wish

to assert that most people have used the word for what is pleasant and for what is desired respectively. . . . My dear sirs, what we want to know from you as ethical teachers, is not how people use a word; it is not even, what kind of actions they approve, which the use of this word good may certainly imply: what we want to know is simply what *is* good. (Moore 1903a, 11–12)

In the end, the strategy consisting in adjusting the definitions to the proof that one is aiming for, leads to losing sight of the question about the nature of *A*—in the example taken by Moore, losing sight of the ethical issue. In brief, confusing characterization and definition makes the expression of a disagreement over the nature of Good impossible. And since confusion does not put an end to the difference of opinions, the disagreement "migrates", so to speak, and manifests itself in "degraded" forms: either in the form of a dispute concerning questions of moral psychology (is what is pleasant identical to what is desirable?), or in the form of a disagreement over linguistic uses (is the word "good" a commonly used substitute for the word "pleasant"?).

The reason why Moore rejects the naturalistic fallacy, i.e., the attempt to prove ethical principle through the use of stipulative definition, is thus that the naturalistic fallacy makes the philosophical enquiry about the nature of Good impossible. In the naturalistic fallacy, definition is used for stopping the philosophical investigation before it starts:

> [When we think we have a definition of "good",] we shall start with the conviction that good must mean so and so, and shall therefore be inclined either to misunderstand our opponent's arguments or to cut them short with the reply, 'This is not an open question: the very meaning of the word decides it; no one can think otherwise except through confusion.' (Moore 1903a, 20–21)

As suggested in the passage, Moore's notorious open question argument should be read from this perspective: that alleged definitions should not be used to stop ethical inquiry is what grounds the fact that the question "is it true that *X* is good?" is never meaningless (i.e., is an open question). The philosopher who falls prey to the naturalistic fallacy (who confuses definition and characterization) considers that the truth of an ethical principle *P* derives from the meaning of the words contained in *P*. There is thus no sense, for him, to denying *P*: "the very meaning of the words decides [the truth of *P*]". It is then the criticism of the naturalistic fallacy that leads Moore to claim that first ethical principles are synthetic and incapable of proof:

> If I am asked, 'What is good?' my answer is that good is good, and that is the end of the matter. Or if I am asked 'How is good to be defined?' my answer is that it cannot be defined, and that is all I have to say about it. But disappointing as these answers may appear, they are of the very last importance. To readers who are familiar with philosophic terminology, I can express their importance by saying that they amount to this: That propositions about the good are all of them synthetic and never analytic; and that is plainly no trivial matter. (Moore 1903a, 6)

The analysis that Moore leads us to in Sections 10–14 is rapid, allusive and, in many ways, dissatisfying. The connection between this and the other development on the naturalistic fallacy one finds in the book is not clear, there are ways out of the fallacy that are not explored, and Moore's conception of analytic truth is no longer ours. But the passage has the virtue of revealing three things that are instructive in the perspective of evaluating Levine's interpretation.

First, there seems to be a tight link between the claim that ethical principles are synthetic and the Principle of Evidence. As we formulated it on page 4, (PoE) implies that philosophical principles "cannot be proved". Yet, as Levine understood it, (PoE) seems to say more than that: it seems to imply that the only way to acknowledge the truth of a principle is to perceive it directly, by the mind's eye, and not through rational argumentation. I am not sure, however, that Moore's notion of intuition should be loaded with such epistemic weight. I will explain

myself more fully about this important issue in Section 5. But for the moment, let me quote Moore's preface:

> I would wish it observed that, when I call such propositions 'Intuitions', I mean merely to assert that they are incapable of proof; I imply nothing whatever as to the manner or origin of our cognition of them. Still less do I imply (as most Intuitionists have done) that any proposition whatever is true, because we cognize it in a particular way or by the exercise of any particular faculty: I hold, on the contrary, that in every way in which it is possible to cognize a true proposition, it is also possible to cognize a false one. (Moore 1903a, x)

Moore makes it clear that intuition does not designate "a manner or origin of our cognition", nor does it designate "a particular way" to know, or the exercise of "any particular faculty". Calling a proposition an intuition is merely a way for him to assert that it is incapable of proof, i.e., synthetic. Thus, when Moore refers to intuition, it seems that he is not opposing two ways of knowing the truth (immediate insight versus discursive argumentation), as Levine suggests. He is, rather, characterizing a proposition as a synthetic truth, i.e., as a proposition whose truth cannot be derived from the meaning of its components. When read in this way, (PoE) does not only imply that philosophical principles are synthetic, (PoE) *just means* that philosophical principles are synthetic. Then, from this perspective, (PoE) appears to be the direct outcome of the rejection of the naturalistic fallacy: to assert that ethical principles are intuitions is just to say that their truth cannot be derived from the meaning of their components.

The second lesson of Sections 10–14 directly follows from the first. There is a hierarchy among the three principles (Aug), (PoA) and (PoE). Far from being an extension of (PoA), (PoE) is in fact at the heart of the argumentation. What Moore attacks is first and foremost the idea that ethical principles are analytic propositions. He wants to reject any attempt to redefine the meaning of the words occurring in these principles so as to derive their truth from those redefinitions. The claim that first principles are synthetic, and the endorsement of (PoE), is the result of this reasoning. Now, viewed from this perspective, (Aug) and (PoA) appear as two subordinate clauses intended to support (PoE). The idea seems to be that if the word "good" were not to refer to one definite thing (Aug) one can be acquainted with (PoA), one would have no reason to oppose the attempt to prove ethical principles. One would not have any means to curb the proliferation of ethical principles, each one grounded on a different stipulative "definition" of *Good*. Why is it not possible to stipulate, as the philosophers who commit the naturalistic fallacy pretend, that good is what is pleasant? Because "good" and "pleasant" refer to two distinct entities (Aug) which are cognitively accessible (PoA). In other words, (Aug) and (PoA) guarantee that there is a basis the hedonist principle should be checked against. They prevent any attempt to adjust the "definition" of the fundamental ethical notions so as to derive our preferred ethical principle.

The third conclusion to be drawn from *Principia Ethica* Sections 10–14, is that Moore's theory of meaning is indeed, as Levine suggests, tightly connected to considerations about method. However, these implications do not resemble at all the ones Levine describes. Indeed, Levine explains that there is an incompatibility between Moore's theory of meaning and the use of conceptual analysis (M1) and argumentation (M2) in philosophical investigation. But a close reading of Moore's criticism of the naturalistic fallacy reveals that Moore targets a position akin to the one Levine attributes to him. Indeed, what Moore attacks is a methodology aiming at shortcutting any rational discussion about ethical principles, exactly like the one Levine ascribes to Moore. Recall the description of Moore's position Levine offers: "if somebody cannot intuit what I have intuited, there is no argument . . . by which I can convince them that what I intuited is really there" (2009, 32). That one should not close a philosophical investigation before it can even get started is precisely the key argument Moore uses to criticize

those philosophers who fall prey of the naturalistic fallacy. Of course, the two stances are not exactly identical. In Levine, a principle is justified by a subjective, incommunicable insight; in Moore's naturalistic fallacy, the justification is provided by the meaning of the names that appear in the statement of the principle. However, the result is the same: both strategies make argumentation over the first principles impossible. When resituated in the context of *Principia Ethica*, Moore's developments on meaning do not seem to lead to (M1) and (M2). I will come back to this point in Section 5.

4. "Old Sidg's" Shadow[14]

In Section 3, I linked Moore's theory of meaning to the criticism of the naturalistic fallacy. In this section I will broaden the focus even more: What is the source of Moore's criticism of the fallacy? Who are the "too many philosophers" that fall prey to the fallacy?

Moore's rejection of the naturalistic mistake is a generalisation of the criticism that Sidgwick addresses to the first utilitarians, Bentham and Mill. The Sidgwickian anchoring of Moore's thought is most visible in Part III of (1903a), devoted to a discussion of the hedonistic principle (H), according to which "pleasure alone is good as an end in itself and for itself" (1903a, 77, 130). Mill and Sidgwick defend (H), Moore discards it. However, in his discussion of Mill and Sidgwick, Moore draws a clear distinction between the criticism he addresses to the former (§§38–44) and the one he addresses to the latter (§§45–57): Mill's justification of (H) is a textbook case of the naturalistic fallacy, while Sidgwick's defence is completely exempt from such a mistake. Better still, it is Sidgwick who first formulated and first criticized the fallacy, in the course of rejecting Mill's attempt of proving (H). Let me briefly explain.

---

[14]On Russell's and Moore's personal relationships with Sidgwick, see Griffin (1989).

In Chapter 4 of *Utilitarianism*, Mill undertakes proving the truth of (H). For this purpose, Mill first defines Good as the object of desire; he then proves that pleasure is the object of desire; by transitivity of identity, that comes down to prove (H). According to Moore, the second step of the proof (that pleasure is the object of desire) is an empirical proposition which belongs rather to moral psychology than to ethics (§42). But it is mostly the first part of Mill's argument that Moore attacks in (1903a, §41): defining good as what is desirable is justified by nothing else than the wish to justify (H). Having postulated what needs to be demonstrated, Mill has no trouble proving what he wants. The so-called proof of hedonism is entirely based on the arbitrary decision that a certain characterization of *Good* is its meaning—it is a paradigmatic example of the naturalistic fallacy.[15] And it is to Sidgwick that Moore attributes the paternity of this criticism. All the way through *The Methods of Ethics*, Sidgwick makes sure to distinguish Mill's justification of (H) from his own. Let me quote for example:

> . . . when we pass from the *adjective* to the *substantive* 'good,' it is at once evident that this latter cannot be understood as equivalent to 'pleasure' or 'happiness' by any persons who affirm— as a significant proposition and not as a mere tautology—that the Pleasure or Happiness of human beings is their Good or Ultimate Good. Such affirmation, which would, I think, be ordinarily made by Hedonists, obviously implies that the *meaning* of the two terms is different, however closely their denotation may coincide. (Sidgwick 1874, 109; see also 387–88, 418ff.)

Sidgwick, just like Moore later on, refuses to consider (H) as a "tautology" (an analytic truth) and stresses that the meanings of "pleasure" and "happiness" are different.

By the end of the first step (§§38–44) of *Principia Ethica* III, Moore, following Sidgwick, has established that (H) is not an

---

[15]Moore (1903a, 65-66): "Mill has made as naïve and artless a use of the naturalistic fallacy as anybody could desire."

analytic truth. But the question of the truth of (H) remains open and constitutes the topic of the second part (§§45–57). On this issue, Moore is opposing Sidgwick:[16]

> So far I have been only occupied with refuting Mill's naturalistic arguments for hedonism; but the doctrine that pleasure alone is desirable may still be true, although Mill's fallacies cannot prove it so. This is the question which we have now to face. The proposition, 'pleasure alone is good or desirable,' belongs undoubtedly to . . . the class of first principles, which are not amenable to direct proof. But in this case, as [Mill] rightly says, 'considerations may be presented capable of determining the intellect either to give or withhold its assent to the doctrine' (p. 7). It is such considerations that Professor Sidgwick presents, and such also that I shall try to present for the opposite view. This proposition that 'pleasure alone is good as an end,' . . . will then appear, in Professor Sidgwick's language, as an object of intuition. I shall try to shew you why my intuition denies it, just as his intuition affirms it. It *may* always be true notwithstanding; neither intuition can *prove* whether it is true or not; I am bound to be satisfied, if I can 'present considerations capable of determining the intellect' to reject it. (Moore 1903a, 74)

Moore's aim, in this second phase, is not to prove that (H) is false (this is impossible since (H) is synthetic), but to present reasons to doubt a proposition whose truth is not logically excluded. These reasons are by no means compelling evidence; they are

"intuitions", i.e., reasons that need to be weighed against the reasons presented by Sidgwick.

I won't detail Sidgwick's and Moore's opposing arguments here. The important point for us is to acknowledge that Sidgwick and Moore do not aim at proving that hedonism is true or false, but are merely presenting fallible considerations 'capable of determining the intellect either to give or withhold its assent'—essentially inferences to the best explanation made in the context of a dialectical confrontation with rival conceptions. In this respect, and despite his opposition to (H), Moore clearly acknowledges his debt to Sidgwick's methodological revolution. It is no exaggeration to say that Moore's criticism of the naturalistic fallacy is just an extension of Sidgwick's opposition to Mill's and Bentham's versions of hedonism.

As Sidgwick's work is not known outside the historians of English-speaking moral philosophy circles, it is important to say a few words on its place within moral philosophy. In his (1874), Sidgwick distinguishes three "methods of ethics", that is, three ways for the human being to rationally justify moral choices. Intuitionism endows the subject with the capacity to immediately perceive his duties; utilitarianism claims that a good action is one that produces the biggest amount of happiness for all; egoism maintains that every individual must promote her own happiness.[17] Behind each of these methods hide different protagonists: Hobbes is the champion of egoism, Whewell and Grote are the representatives of intuitionism, Mill and Bentham are the spokesmen for utilitarianism. By simplifying, one could sum up Sidgwick's project by saying that it consists, against egoism, in building a bridge between utilitarianism and intuitionism. Indeed, for Sidgwick, the main dividing line is not the one opposing utilitarianism and intuitionism (as it was supposed to be the case at the time), but between the one opposing

---

[16]See also Moore (1903a, 63–64) and (1903a, 77):

Well, then, we now proceed to discuss Intuitionistic Hedonism. And the beginning of this discussion marks, it is to be observed, a turning-point in my ethical method. . . . We are [now] coming to the question, for the sake of answering which Ethics exists, the question what things or qualities are good. Of any answer to this question no direct proof is possible, and that, just because of our former answer, as to the meaning of good, direct proof was possible. We are now confined to the hope of what Mill calls indirect proof, the hope of determining one another's intellect; and we are now so confined, just because, in the matter of the former question we are not so confined. Here, then, is an intuition to be submitted to our verdict—the intuition that pleasure alone is good as an end—good in and for itself.

[17]The utilitarian, just like the partisans of egoism, are hedonists: both adhere to the same principle (H).

the two forms of hedonism (1874, 85–88).[18] More precisely, Sidgwick considers that Bentham and Mill have artificially accused the opposition to intuitionism, by presenting utilitarian ethics as a genuine alternative to the morality of common sense. And this last point is linked to Sidgwick's methodological innovation Moore praised so much: it is because they considered (H) as an analytic truth, that Bentham and Mill claimed that nothing (no reference to moral intuition or to common sense) could run counter to the application of the utilitarian calculus, and it is because he considered that (H) cannot be justified without resorting to intuition, that Sidgwick was opposed to Bentham's and Mill's "aggressive" and "negative" understanding of utilitarian ethics.[19] It is then important to understand that Sidgwick's "methodological" claims were connected to more "substantial" issues.[20]

To summarize: with Sidgwick, the idea that the fundamental ethical principles are not analytic truths leads to a new argumentative strategy based on the dialectical confrontation with rival conceptions, in which the truth of ethical principles is not derived from a proof, but from a sort of inference to the best explanation. When Sidgwick called (H) an intuition, he did not mean to suggest that the truth of (H) could be perceived by the mind's eye in a sort of instantaneous flash. He just marked his opposition to a dogmatic version of hedonism that considered (H) as an analytic truth.

According to Rawls, Sidgwick's break with the old versions of utilitarianism represents a major turning-point in the history of moral philosophy:[21]

The Methods of Ethics . . . is the first truly academic work in moral philosophy . . . , modern both in its method and in the spirit of its approach. It treats moral philosophy as any other branch of knowledge. It undertakes to provide a systematic comparative study of moral conceptions, starting with those which historically and by present assessment are the most significant. . . . Sidgwick's originality lies in his conception of the subject of moral philosophy, and in his view that a reasoned and satisfactory justification of any particular moral conception must proceed from a full knowledge and systematic comparison of the more significant moral conceptions in the philosophical tradition. (Rawls 2000, 376)

Note that in Rawls' assessment, Sidgwick is praised for having introduced scientific method in moral philosophy. For him, it is obvious that Sidgwick's agreement with intuitionism should not be seen as a defence of a subjective and introspection-based method, but, on the contrary, as a way to ground the justification of any ethical principle on dialectical arguments (on "a full knowledge and systematic comparison of the more significant moral conceptions in the philosophical tradition").

This detour by Sidgwick shows then how Moore's naturalistic fallacy argument is deeply rooted in Sidgwick's methodological breakthrough. It also shows the importance and the significance

---

[18]Sidgwick (1874, 85): "It seems to me undeniable that the practical affinity between utilitarianism and intuitionism is really much greater than that between the two forms of hedonism."

[19]For a presentation of the "aggressive" and "negative" version of utilitarianism, see Sidgwick (1874, 423).

[20]Thus, the opposition between Sidgwick, on the one hand, and Bentham and Mill, on the other, is not only conceptual and methodological, but also political—Sidgwick is a reformist and not a revolutionary. On this aspect, see Schneewind (1977, 129–34, 180–82, 260–65).

[21]Rawls (1999, 44; see also 507):

I wish to stress that in its initial stages at least a theory of justice is precisely that, namely, a theory. . . . A theory of justice is subject to the same rules of method as other theories. Definitions and analyses of meaning do not have a special place: definition is but one device used in setting up the general structure of theory. Once the whole framework is worked out, definitions have no distinct status and stand or fall with the theory itself. In any case, it is obviously impossible to develop a substantive theory of justice founded solely on truths of logic and definition. . . . There is no other way to give an account of our considered judgments in reflective equilibrium. This is the conception of the subject adopted by most classical British writers through Sidgwick. I see no reason to depart from it.

of this advance: Sidgwick cannot be portrayed as an "intuitionist" in Levine's sense.

## 5. The Methodological Import of Moore's Semantic Framework

Let me now come back to Levine's interpretation. Levine claims that Moore's theory of meaning is the source of a radical philosophical "intuitionism", according to which no clarification of concepts (M1) and no reasoned justification of the first principles (M2) is possible. The analysis of the two previous sections leads us to reviewing this conclusion.

The main point of disagreement between Levine's interpretation and mine concerns the place of (PoE) in Moore's theory. Levine does not formulate this principle. For him, all happens as if the intuition which informs the mind about the meaning would also inform it about the truth of the first philosophical principles. Yet it turns out that the target of Moore's (and Sidgwick's) reasoning is Bentham and Mill, and that the claim that ethical principles cannot be proved (PoE) plays an instrumental role in this criticism. It is then (PoE) that is the rationale behind (Aug) and (PoA). These principles are only subordinate clauses intended to back (PoE) against any attempt to stipulate that a term has a certain meaning for the purpose of proving the truth of a certain principle. Of course, (PoE) cannot be derived from (Aug) and (PoA) alone. But (Aug) and (PoA) provide reasons for claiming that ethical principles are synthetic, i.e., to block any attempt to derive their truth from a stipulative definition.

This emphasis put on (PoE) leads me to object to Levine's idea that Moore's semantical framework leads to (M2). Settling philosophical disputes about first principles by resorting to stipulative definition is not the same thing as doing so by resorting to subjective intuition. However, both strategies lead to the same result: banning rational argumentation from philosophical enquiry (M2). Now, we have seen that the reason Moore refused the first strategy was precisely that it led to (M2). It seems then impossible to attribute to Moore the second strategy, since it would lead him to endorse the conclusion he is explicitly rejecting in his development on the naturalistic fallacy! I am not claiming here that Moore actually succeeds in meeting his goal, that is, I am not claiming that Moore actually secured the possibility of using rational argumentation in philosophical investigation. My point is only that this was a part of his methodological agenda; and then, that (unless we imagine that Moore was flatly incoherent) Levine's claim that there is an incompatibility between Moore's semantical framework and rational argumentation cannot be upheld.

Here is a schema (S2) which illustrates the relations between (Aug), (PoA), (PoE) and (M2) as I see it:
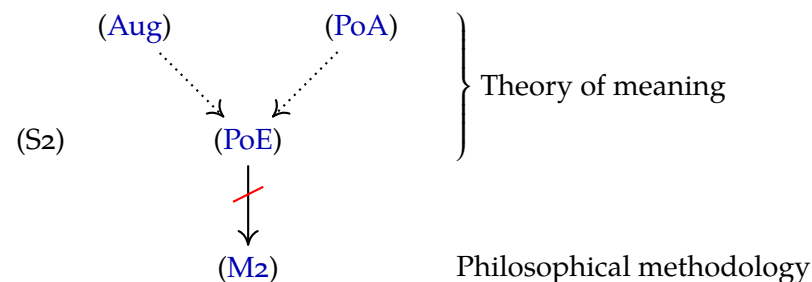
The solid arrow represents the implication relation. The red strike indicates that the implication does not hold (Moore's theory of meaning is not incompatible with rational argumentation). The dotted arrow represents "weak" implication: (Aug) and (PoA) are reasons that intervene in the justification of (PoE), but (PoE) does not follow deductively from (Aug) and (PoA) alone.

One question remains, however. I said that Moore's theory of meaning is compatible with (M2); but what about (M1)? Doesn't Moore's theory of meaning (and notably (PoA)) doom any analysis to be merely making explicit of what is already known?

Two meanings of analysis need to be distinguished: analysis as a process of decomposition of a complex entity, and analysis as a process of distinction of a term in relation to other related terms. In the first sense, it is perfectly clear that, in Moore (1903a), analysis is just making something explicit. Thus, for instance, there is no analysis possible of the simple term *Good*. And when the term is complex and decomposition is possible, then analysis is not informative since acquaintance with the term presupposes acquaintance with all its constituents. But the situation drastically changes once the expression "analysis" designates the process of distinguishing a meaning from other related meanings.

In *Some Main Problems of Philosophy*, published in 1953, but written around 1910, Moore explains:

> To say that it is difficult to be sure what [the expressions 'really exist', 'real', 'is true'] mean, which is, I think, true, does not therefore imply that we cannot perfectly easily understand the meaning of sentences in which they occur, and be quite certain that some such sentences are true and others false. We can do both these things, without, in a sense, knowing exactly what these words mean. For by knowing what they mean is often meant not merely understanding sentences in which they occur, but being able to *analyze* them, or knowing certain truths about them—knowing, for instance, exactly how the notions which they convey are related to or distinguished from other notions. And obviously we may be quite familiar with a notion itself, it may be quite readily conveyed to us by a word, even though we cannot *analyze* it or say exactly how it is related to or distinguished from other notions. (Moore 1953, 205, see also 267, 269, 307; my emphasis)

The process of distinguishing a concept from other concepts is here named "analysis". And as specified by Moore in the passage, analysis in this sense is clearly informative. Someone can understand an expression without being able to specify how its meaning is linked to, and distinguished from, other meanings that are close to it. Someone acquainted with a meaning does not know all the truths about this meaning. It could even be said that (PoE) leads to consider any analysis (in the sense here discussed) as informative: grasping the meanings of "pleasure" and "good" (for example) does not determine in itself whether they are identical or different, i.e., the truth-value of (H). (On this see Wishon 2017 and below.)

If we follow this lead, we can then resist Levine's claim that (Aug) and (PoA) imply (M1). In Moore, analysis is not viewed merely as making explicit a content that is already present, but as the exploration of one term's relations with other similar terms—relations which, being external, cannot be derived from the meaning of the terms. Levine uses a very restricted notion of analysis, according to which, for instance, there is no analysis of Good (Good being a simple term). But the investigation of the relations between the concept Good and the meaning with which it is usually associated (pleasure, desirability, etc.), which is the topic of (1903a), can be counted as an analysis in the extended sense described in (1953). What Moore said there can be applied word for word to this case: "we may be quite familiar with [the notion of Good], it may be quite readily conveyed to us by a word, even though we cannot analyse it or say exactly how it is related to or distinguished from other notions".

I can then complete the pattern (S2) with a new pattern (S3), in which the two methodological implications that Levine draws from Moore's theory of meaning are rejected:
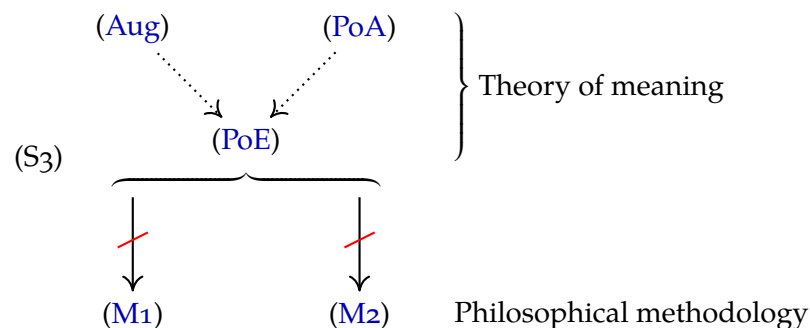
My interpretation of Moore's semantical framework leads then to conclusions that are opposed to those drawn by Levine. Where does this discrepancy come from?

Levine considers Moore's intuition as an epistemic term: intuition (or acquaintance) would be, for Moore, a source of knowledge. Levine's view is first and foremost based on *The Refutation of Idealism*, where free rein is given to epistemological considerations;[22] and I agree that Levine's epistemic interpretation of intuition is difficult to resist when due account is taken of (1903b). But as I say at the end of Section 3, Moore seems willing to leave out this conception of intuition in (1903a). In the preface, Moore makes it clear that intuition does not designate "a manner or origin of our cognition". The accent is put, instead, on the gap between understanding a proposition and recognizing its truth: a proposition is called an intuition when its truth does not derive from the meaning of its constituents (when the proposition is "incapable of proof" or "synthetic"). By (Aug) and (PoA), understanding the proposition P ("good is pleasure") requires acquaintance with the elements (*Good*, *Pleasure*) of P; but being acquainted with the constituents of P is not enough (by (PoE)) to assess the truth of P: one additional thing, called intuition, is needed. Moore's notion of intuition of a truth is thus contrasted, in *Principia Ethica*, with the notion of acquaintance with a non-propositional thing.

In the interpretation I presented above, I deepen this divide—a divide that Levine's epistemic reading tends to erase. Indeed, even if he does not deny the difference between the propositional and the non-propositional cases, Levine seems to glide from the intuitive understanding of terms to the intuitive apprehension of truths: what is important for him is that, in both cases, intuition is a source of knowledge. I consider, on the contrary, that there is no single type of knowledge that intuition is supposed to allot. Acquaintance with non-propositional meanings cannot give us any clue about the truth of a proposition.[23] This is, in substance, what (PoE) amounts to. The "typological" gap between non-propositional and propositional knowledge is the reason why a reference to a different kind of epistemic device is needed to account for the truth and falsity of first principles.

This position echoes the emphasis put by some recent scholars on the contrast made by Russell between knowledge of a thing and knowledge of a truth in Chapter 4 of *The Problems of Philosophy*.[24] In his (2015), Ian Proops has insisted on the fact that, in 1912 at least, Russell held the thesis that knowledge by acquaintance is theoretically independent of knowledge of

---

[22]See Levine (2009, 19–20):

For Moore, "sensation and thought" are "both forms of consciousness or, to use a term that seems to be more in fashion just now, they are both ways of experiencing" (1903b., 7), so that "the nature of that peculiar relation which I have called 'awareness of anything'... is involved equally in the analysis of *every* experience—from the merest sensation to the most developed perception or reflexion" (1903b, 29). For Moore, to have an "idea" of an entity—whether by sensation or by thought—is to be "aware of" or to "know" that entity, an entity which is not (in general) a mental item.

Also see Baldwin (1990, 12–16, 49–52).

[23]And this is so even if a proposition is thought by Moore as a sort of complex entity. On Moore's early theory of propositions and its difficulties, see Moore (1953, 252–69) and Baldwin (1990, 161ff.).

[24]Russell (1912, 44):

The word 'know' is... used in two different senses. (1) In its first use it is applicable to the sort of knowledge which is opposed to error, the sense in which what we know is *true*, the sense which applies to our beliefs and convictions, i.e. to what are called *judgements*. In this sense of the word we know *that* something is the case. This sort of knowledge may be described as knowledge of *truths*. (2) In the second use of the word 'know' above, the word applies to our knowledge of *things*, which we may call *acquaintance*. This is the sense in which we know sense-data.

truth.[25] Travelling in the same circle, Donovan Wishon ([2017](#), 365) holds the view that acquaintance does not safeguard us from misidentifying the objects of our acquaintance. To be acquainted with an object *A* is not to have a "discriminating conception" allowing us to propositionally know that *A* is distinct from another object *B*.[26] What is common between these views and my own interpretation is the belief that it is crucial not to align the knowledge of a truth (what Moore calls intuition here) with the knowledge of a non-propositional entity (what one could label acquaintance).[27] While Levine models intuition (i.e., propositional knowledge-that) on acquaintance (i.e., non-propositional knowledge-of), I claim that Moore views the former as a heterogeneous epistemic device, which works in a different way than, and is independent from, the latter. My disagreement with Levine seems then to come from our different appreciation of the relation between what Russell, in 1912, calls direct knowledge-of and direct knowledge-that.

---

[25]Proops ([2015](#), 796): "Knowledge of truth . . . is one kind of epistemic state, and knowledge of things by acquaintance quite another—and 'theoretically' one may have knowledge of a thing by acquaintance without knowing a single truth about it." See Russell ([1912](#), 73): "The particular shade of colour that I am seeing may have many things said about it, I may say that it is brown, that it is rather dark, and so on. But such statements, though they make me know truths *about* the colour, do not make me know the colour itself any better than I did before."

[26]As Baldwin pointed out, Moore seems to make room for the idea that one can be mistaken about the identity of the object one is acquainted with. In his ([1990](#), 63–64), Baldwin quotes a passage of Moore ([1901–02](#)), in which Moore opposes McTaggart on "the following very elementary psychological fact" ([1901–02](#), 202): "A man may be fully convinced, in any ordinary sense of conviction . . . that a certain predicate does not attach to a certain subject, and yet whenever he imagines that subject . . . may quite unconsciously include in it the very predicate of which both then and at all other times he is ready to asseverate the absence."

[27]Proops considers James to be the source of Russell's 1912 distinction between knowledge of a thing and knowledge of a truth. The present paper suggests that the idea is present in Moore and Russell as early as 1903—that it lies at the heart of Moore's criticism of the naturalistic fallacy.

6. Moore and Russell: Mathematical Definition and Denoting Concepts

Until now, I have targeted Levine's reading of Moore, and have said hardly anything about Russell. This is in line with what I announced in my introduction: the goal of the paper is to show that informative analysis and rational argumentation are compatible with Moore's semantical framework. But still, the question naturally arises: What lessons about Russell can be learned from what I have said so far? Already in 1903, there are important differences between Moore and Russell. For instance, as I pointed out in Section [2](#), Moore did not accept the class definition of number, Russell did. Levine's interpretation has the merit of explaining this disagreement: for in making room for informative analysis and rational argumentation, Russell, somewhere before the publication of *The Principles* (probably in June 1902; see Levine [2009](#), 45–46, 77) would be led to give up (Aug). As I don't see any contradiction between Moore's semantic and Peano-inspired analyses, I can't agree with this explanation. But how do I account, then, for the difference between Russell's and Moore's attitude toward (Num)? In this last section, I would like to outline an answer based on Russell's theory of denoting concepts. I do so because, as we have begun to see in Section [3](#), Russell's theory of denoting concepts has natural connections with Moore's considerations about meaning, definition and analysis. But I do so, as well, in order to show that Levine neglects certain alternatives that seem to be at least as plausible as the one he provides.

The core of Russell's theory of denoting concepts is expressed in the following sentence: "a concept *denotes* when, if it occurs in a proposition, the proposition is not *about* the concept, but about a term connected in a certain particular way with the concept" ([1903a](#), 53).[28] In Chapter 5 of ([1903a](#)), Russell devotes a great

---

[28]Russell continues ([1903a](#), 53): "If I say 'I met a man', the proposition is not

deal of space to the attempt to characterize the different kinds of denoting concepts. To save time, I will focus here only on the case where the denotation is a single entity. Russell explains that, starting from "a class-concept [C] of which there is only one instance" (1903a, 62), it is possible to form the concept *the C* which denotes "one single definite term" (1903a, 62). Let's say that *E* is the unique entity denoted by *the C*; and consider the two propositions $P_1 =$ *E is F*, and $P_2 =$ *the C is F*. Since they have different constituents, P_1 and P_2 are different propositions. Russell, however, claims that P_1 and P_2 are about exactly the very same entity, namely *E*. That is, Russell introduces a distinction between the entities which occur in a proposition (*E* and *the C*, respectively), and the entities the proposition is about (*E*). Gideon Makin calls aboutness-shifting "the property of denoting concepts of making the propositions in which they occur be about an entity other than themselves" (2000, 17–18). That certain entities enjoy the property of aboutness-shifting is the very essence of Russell's view.

There seems to be an obvious connection between Moore's criticism of the naturalistic fallacy, as I have presented it in Section 3, and Russell's view on denoting. Indeed, what I called univocal characterization seems to correspond to *the*-denoting concepts. However, Moore never accepted aboutness-shifting. Speaking about Russell's theory, he wrote in his unpublished review of *The Principles*:[29]

---

about *a man*: this is a concept which does not walk the streets, but lives in the shadowy limbo of the logic-books. What I met was a thing, not a concept, an actual man with a tailor and a bank-account or a public-house and a drunken wife."

  [29]Here is the whole passage:

Mr. Russell points out that there is a difficulty in discovering the exact meaning of the . . . class of proposition that we express by such phrases like '*All* men are mortal', 'if *any* man is mortal', 'I met *a* man', '*The* man who did this'. It is plain that in such cases we do not generally mean to assert a simple relation between two predicates, as is assumed by logicians

Mr. Russell [holds] that the thing about which a proposition is, may not be a constituent in that proposition. But this conclusion is undoubtedly wrong. (Moore 1905–06, 27)

For Moore, two propositions P_1 and P_2 that have different subjects (*E* and *the C*) can't be both about the same thing. The argument seems to be this: if P_2 were to be about *E*, then the phrase "the C" would be both about *the C* and about *E*. Now, by (Aug), "E" and "the C" designate two different things. Thus aboutness-shifting seems to go against (Aug), and should then be given up.

I think however that the situation is more complicated than that. First, it is not true that aboutness-shifting goes against (Aug). In his discussion of Russell's 1903 theory, Makin (2000, 18–20) emphasizes that aboutness-shifting is not a linguistic, but a logical property: the relation between the meaning and the denotation of a denoting concept must be carefully distinguished from cases of lexical ambiguity. Thus, for Russell, the phrase "the C" designates a unique entity, namely *the C*; and it is this entity which subsequently shifts the aboutness of the proposition in which it occurs towards *E*. Russell still endorses (Aug): the denoting phrase "the C" is not ambiguous in its meaning at all. There is a second complication. Moore's refusal to distinguish between occurrence and aboutness has a drawback. Pushed to its extreme, it amounts to regarding any identity statement as false. Consider the identity $P_3 =$ *E is the C*. If *E* and *the C* are different entities, and if P_3 is about *E*

---

who take an *intensional* view; . . . we undoubtedly do mean to assert that the two properties are related *in respect of their extension* . . . . But on the other hand it is equally plain that, when we make such a proposition, we by no means always have the whole extension, about which we seem to assert something, before our minds. It would seem therefore, at first sight, as if could be cognisant of a proposition *about* a thing, without having that thing before our minds. And in his book, Mr. Russell actually adopts this view; and even adds the conclusion, that the thing about which a proposition is, may not be a constituent in that proposition. But this conclusion is undoubtedly wrong. (Moore 1905–06, 26–27)

and *the C*, as required by Moore, then P_3 cannot be true: *E* is a different meaning than *the C*. This is clearly a difficulty for Moore. All the naturalistic fallacy discussion in Sections 10–14 of (1903a) implies that an entity *E* can be univocally characterized by a phrase that does not mean *E*. Moore seems to routinely accept that the phrase "what is caused by stimulating the normal eye by such luminous vibration" is, in some sense, about the colour *yellow*. However, owing to his rejection of aboutness-shifting, Moore cannot even explain how "the C" can be about *E*; for him, "the C" is always about *the C*.

For these two reasons, and despite Moore's opposition, I think we should consider Russell's theory of denoting concepts as an extension of Moore's semantical framework. All in all, Russell agrees with Moore that, in P_3, the meaning of "the C" is different from the meaning of "E". He thus agrees with the criticism of the naturalistic fallacy. But he can make room for the idea that P_3 is true, and that P_1 and P_2 are about the very same entity.

Now, my claim is that Russell's new theory can explain the difference between Russell's and Moore's attitudes toward (Num). Indeed, Russell's theory gives rise to a new conception of definition,[30] which makes a distinction between philosophical and mathematical definitions:

> It is necessary to realize that definition, in mathematics, does not mean, as in philosophy, an analysis of the idea to be defined into constituent ideas.... Mathematical definition consists in pointing out a fixed relation to a fixed term, of which one term only is capable; this term is then defined by means of the fixed relation and the fixed term. (Russell 1903a, 27)

Philosophical definitions correspond to what Moore calls definitions, i.e., to the analyses of complexes into their constituent terms. Mathematical definitions correspond to what we have called univocal characterizations, which had no clear status in Moore. In Russell, aboutness-shifting explains why, in a mathematical definition, the *definiens* is about the *definiendum*, even though the two phrases do not have the same meaning. This account explains how definitions can be informative:

> It is a curious paradox, puzzling to the symbolic mind, that definitions, theoretically, are nothing but statements of symbolic abbreviations, irrelevant to the reasoning and inserted only for practical convenience, while yet, in the development of a subject, they always require a very large amount of thought, and often embody some of the greatest achievements of analysis. This fact seems to be explained by the theory of denoting. An object may be present to the mind, without our knowing any concept of which the said object is *the* instance; and the discovery of such a concept is not a mere improvement in notation. The reason why this appears to be the case is that, as soon as the definition is found, it becomes wholly unnecessary to the reasoning to remember the actual object defined, since only concepts are relevant to our deductions. In the moment of discovery, the definition is seen to be *true*, because the object to be defined was already in our thoughts; but as part of our reasoning it is not true, but merely symbolic, since what the reasoning requires is not that is should deal with *that* object, but merely that it should deal with the object denoted by the definition. (Russell 1903a, 63)

Take P_3 again. P_3 is not an analytic truth. If P_3 is true, it is because *E* is the only instance falling under the concept *C*; and this is, as Russell says, a discovery. But once the truth of P_3 is established, "it becomes wholly unnecessary to the reasoning to remember [*E*] since only concepts [*the C*] are relevant to our deductions". I take this phrase to mean that, reasoning about *E*, we can use "the C" in our deductions, since *the C* denotes *E*. Once again, it is crucial to realize that Russell does not commit the naturalistic fallacy: in a mathematical definition, one does not stipulate that the meaning of "E" is the same as the meaning of "the C". P_3 is a synthetic truth, that should

[30]See Russell (1903a, 62–63): "It is owing to this notion [*the*-denoting concepts] that mathematics can give definitions of terms which are not concepts—a possibility which illustrates the difference between mathematical and philosophical definition."

be justified by an argument showing that *E* is the sole entity satisfying *C*. But at the same time, P_3 is a genuine definition, in that we can forget the meaning of "E", and use "the C" instead in all our deductions.

Let me come back to (Num). Owing to the preceding, one could claim that (Num) is, for Russell, a mathematical definition. This would explain Russell and Moore's divergence in a more economical way than in Levine: both Moore and Russell agree that the *definiens* and the *definiendum* do not have the same meaning, but Russell, against Moore, can account for the fact that "the number of α" and "the class of classes similar to α" are about the same thing. This explanation raises two problems, however. The first one, easy to overcome, is that what is defined in (Num) is not a single entity, but a class. For this reason, mathematical definition does not seem to be applicable to (Num). Russell is aware of this difference, but, as the end of §63 shows, he does not attach much significance to it: "the notion of *the* is always relevant in definitions".[31] The second problem, which has been raised by Levine, is more serious. Levine recognizes that there is a stage in Russell's thought where mathematical definitions were contrasted with philosophical definitions. But he maintains that this stage is over when (Num) is, in June 1902, finally accepted. From this date onwards, (Num) is philosophically and mathematically adequate: according to Levine, the two phrases "1" and "the class of all singletons" are not only about the same thing, they have the same meaning (Levine 1998, 115–6; 2009, 45–46). Or more exactly, (Aug) is now given up, the meaning of "1" is vague, and the set-theoretical definition should be

regarded as a precisification of a vague notion. In any case, the theory of denoting no longer plays any role in Russell's thought.

I disagree with Levine's analysis. First, textual evidence shows that Russell regards the theory of denoting concepts as a central feature of definition long after 1902. For instance, at the end of Russell (1903b, 307), written during the second half of 1903, Russell writes: "It is chiefly for the sake of Dfs that denoting is required". Moreover, there is at least one passage, dating from 13 June 1904 (see Schmid 2001, 407), where Russell explains that vagueness terminology pertains to psychological discourse, while denoting terminology pertains to logic. It seems, thus, that Levine exaggerates the gap between the denoting-based and the vagueness-based accounts of definition. Secondly all the passages of (1903a) that Levine quotes for supporting the idea that (Num) is accepted as a genuine definition (essentially 1903a, 111–12, 115–16) could be understood as passages promoting the philosophical role of mathematical definition. Russell would not claim that the meaning of "E" is the same as the meaning of "the C", but that the meaning issue is perhaps not as crucial as Moore believed it to be.[32] Finally, the vagueness view that Levine attributes to Russell in 1903 seems incompatible with the way Russell deals with definitions. According to Levine, a definition is for the Peanian Russell a sort of analytic stipulation which sets limits, somewhat arbitrarily, to what has no definite boundaries.[33] If such were the case, then Russell

---

[31]Russell (1903a, 63): "In most actual definitions of mathematics, what is defined is a *class* of entities, and the notion of *the* does not then explicitly appear. But even in this case, what is really defined is *the* class satisfying certain conditions . . . . Thus the notion of *the* is always relevant in definitions; and we may observe generally that the adequacy of concepts to deal with things is wholly dependent upon the unambiguous denoting of a single term which this notion gives."

[32]For instance, let me quote the end of (1903a, 115–16): "philosophically we may admit that every collection of similar classes has some common predicate applicable to no entities except the classes in question, and if we can find, by inspection, that there is a certain class of such common predicates, of which one and only one applies to each collection of similar classes, then we may, if we see fit, call this particular class of predicates the class of numbers. For my part, I do not know whether there is any such class of predicates, and I do know that, if there be such class, it is wholly irrelevant to Mathematics." This does not sound to me like a denial of the difference between philosophical and mathematical definition.

[33]As a matter of fact, the vagueness view that Levine attributes to the

should have denied that definitions are, from an epistemic point of view, discoveries—which he did not in §63. More generally, if such were the case, then Russell should have espoused a sort of Hilbertian view, according to which the mathematician is free to define the mathematical concepts as he or she wants. This is at odds with what Russell did in 1903, where he always sought to justify his different definitions of mathematical concepts.[34]

For these reasons, I conclude that the denoting-based theory of mathematical definition can account for the difference between Moore's and Russell's attitude *vis-à-vis* the class-definition of number. Of course, as it is well-known, Russell will finally abandon the theory of denoting concepts in 1905. It may well be that the reasons Russell has for doing so (which are still very much discussed) are not foreign to the difficulties that Levine points out. But even plagued with its problems, the theory of denoting concepts, it seems to me, can be conceived as a way to accommodate the Moorean semantical frameworks to the requirements of Peanian analytical practices.

7. Conclusion

I have been seeking to highlight the existence of a line that could be called "Sidgwickian" in Moore and Russell's thought. Rejecting the old utilitarian attempts to prove, by means of definition,

---

Peanian Russell is close to the view Moore attacked in *Principia Ethica* under the naturalistic fallacy banner.

[34]Most of the time, Russell did not give only one definition of a given mathematical concept, but several possible ("formally correct") ones; he then attempted to justify his choice by balancing the pros and the cons of the various available options. See, for instance, the definition of open order (that Russell calls betweenness) in *The Principles*, Chapter 24. Three different conceptions, all "formally correct", are first presented; after having compared them, Russell chooses to say that y is between x and z means that there is a certain asymmetric and transitive relation R such that xRy and yRz. For more examples of adjudication between different formally correct analyses, see Gandon (2012, chap. 7).

fundamental ethical principles leads to adopting (PoE). And it is to secure (PoE) that Moore came to affirm that the meaning of words is unequivocally set (Aug) and directly accessible to the mind (PoA). My thesis is thus that what Levine calls Moore's theory of meaning is just a prolongation of the "Sidgwickian" gesture that consists of breaking off from the "stipulative" mode of justifying ethical principles which prevailed in Mill and Bentham. From this, it follows that subscribing to (PoE), (PoA) and (Aug) does not lead Moore to adopting an insight-based methodology, which excludes conceptual analysis (M1) and rational argumentation (M2). On the contrary, when read in the light of the naturalistic fallacy discussion, Moore's theory of meaning appears to be an attempt to restore the place of conceptual analysis and rational argumentation in philosophy. In claiming that Moore's semantic framework is compatible with a "Peanian" methodology, I did not mean however to mitigate the differences between Russell and Moore. The goal was rather to frame a conceptual space in which these differences could be better located.

Russell is often criticized for having espoused a "visual" conception of understanding, according to which what one is thinking is always fully understood, and according to which no analysis can ever provide one with a better understanding.[35] Levine's

---

[35]See for instance Burge (1990, 59):

From the beginning, Russell allowed little or no room for the idea that one could think with notions that one only partially understands. Russell assimilated understanding to a non-propositional, vision-like conception of knowledge, called acquaintance, which propositional abilities were supposed to presuppose. . . . Encouraged by his interest in establishing logicism, Russell did emphasize that through analysis one could enlarge one's powers of understanding and gain insight into the foundations of notions that one is already thinking with. But because he treated understanding in terms of acquaintance rather than in terms of the use of a theory, Russell never allowed this emphasis to threaten the principle of acquaintance . . . . Thus the implicit tension between the principle of acquaintance and the natural view that with 'analysis' we obtain a better understanding of the *analysanda* never comes to a head in Russell's work.

goal is to debunk this kind of interpretation. I agree with Levine that it is both worthwhile and urgent to put an end to this understanding of Russell's thought. But to salvage Russell the Peanian, Levine puts all the sins upon Moore and Russell the Moorean. In this paper, I have suggested that Moore and Russell's "visual" model of understanding is more complicated and flexible than what appears at first, and that Levine's "scapegoating" strategy is not justified. This move leads me to defend a picture of Russell's evolution in which local continuities play a more important role than in Levine's view.

Acknowledgements

Sébastien Gandon
Université Clermont Auvergne
sgandon0@gmail.com

References

Baldwin, Thomas, 1990. *G. E. Moore*. London: Routledge.

Burge, Tyler, 1990. "Frege on Sense and Linguistic Meaning." In *The Analytic Tradition*, edited by D. Bell and N. Cooper, pp. 30–60. Oxford: Blackwell.

Gandon, Sébastien, 2012. *Russell's Unknown Logicism*. Basingstoke: Palgrave Macmillan.

Griffin, Nicholas, 1989. "Russell and Sidgwick." *Russell* 9: 12–25.

Hylton, Peter, 1990. *Russell, Idealism and the Emergence of Analytic Philosophy*. Oxford: Clarendon Press.

Levine, James, 1998. "From Absolute Idealism to *The Principles of Mathematics*." *International Journal of Philosophical Studies* 6: 87–127.

——, 2007. "Analysis and Abstraction Principles in Russell and Frege." In *The Analytic Turn*, edited by M. Beaney, pp. 51–74. New York and London: Routledge.

——, 2009. "From Moore to Peano to Watson: The Mathematical Roots of Russell's Naturalism and Behaviorism." *The Baltic International Yearbook of Cognition, Logic and Communication* 4: 1–126.

——, 2016. "The Place of Vagueness in Russell's Philosophical Development." In *Early Analytic Philosophy: New Perspectives on the Tradition*, edited by S. Costreie, pp. 161–212. New York: Springer.

Makin, Gideon, 2000. *The Metaphysicians of Meaning: Russell and Frege on Sense and Denotation*. London and New York: Routledge.

Mill, J. S., 1861. *Utilitarianism*. Reprinted in *The Collected Works of John Stuart Mill*, vol. 10, *Essays on Ethics, Religion and Society*, edited by John Robson. Toronto: University of Toronto Press, 1969.

Moore, G. E., 1901–1902. "Mr. McTaggart's 'Studies in Hegelian Cosmology'." *Proceedings of the Aristotelian Society* n. s. 2: 177–214.

———, 1903a. *Principia Ethica*. Cambridge: Cambridge University Press.

———, 1903b. "The Refutation of Idealism." Reprinted in *Philosophical Studies*, pp. 1–30. London: Routledge & Kegan Paul, 1922.

———, 1905–06. "Russell's *Principles of Mathematics*." Unpublished typescript. Bertrand Russell Archives, McMaster University.

———, 1942. "A Reply to My Critics." In *The Philosophy of G. E. Moore*, edited by P. A. Schilpp, pp. 535–677. Evanston: Northwestern University Press.

———, 1953. *Some Main Problems of Philosophy*. London: George Allen & Unwin.

Proops, Ian, 2014. "Russellian Acquaintance Revisited." *Journal of the History of Philosophy* 52: 779–811.

———, 2015. "Certainty, Error, and Acquaintance in *The Problems of Philosophy*." In *Acquaintance, Knowledge, and Logic: New Essays on Bertrand Russell's* The Problems of Philosophy, edited by D. Wishon and B. Linsky, pp. 45–63. Stanford: CSLI Publications.

Rawls, John, 1999. *A Theory of Justice*, 2nd ed. Cambridge, MA: Harvard University Press. (First edition 1971.)

———, 2000. *Lectures on the History of Political Philosophy*, edited by S. Freeman. Cambridge, MA: Harvard University Press.

Russell, Bertrand, 1900. *A Critical Exposition of the Philosophy of Leibniz*. Cambridge: Cambridge University Press.

———, 1903a. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

———, 1903b. "On Meaning and Denotation." In *The Collected Papers of Bertrand Russell*, vol. 4, *Foundations of Logic 1903–05*, edited by A. Urquhart, pp. 314–58. London: Routledge, 1994.

———, 1907. "The Regressive Method of Discovering the Premises of Mathematics." Reprinted in *Essays in Analysis*, edited by Douglas Lackey, pp. 272–83. London: George Allen & Unwin, 1973.

———, 1912. *The Problems of Philosophy*. London: Williams and Norgate.

———, 1992. *The Theory of Knowledge: The 1913 Manuscript*, edited by E. R. Eames and K. Blackwell. London: Routledge.

Schmid, A. F., 2001. *Bertrand Russell / Louis Couturat (1897–1913). Correspondance sur la philosophie, la logique et la politique*. Paris: Kimé.

Schneewind, J. B., 1977. *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford: Clarendon Press.

Sidgwick, Henry, 1874. *The Methods of Ethics*. London: Macmillan.

Wishon, Donovan, 2017. "Russellian Acquaintance and Frege's Puzzle." *Mind* 126: 321–37.